

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

## HIV-2 nucleotide sequences

Patent Number: EP0750041

Publication date: 1996-12-27

Inventor(s): ALIZON MARC (FR); SONIGO PIERRE (FR); GUETARD DENISE (FR); MONTAGNIER LUC (FR); TIOLLAIS PIERRE (FR); CHAKRABARTI LISA (FR); GUYADER MIREILLE (FR); DESROSIERS RONALD (US); CLAVEL FRANCOISE (US)

Applicant(s):: CENTRE NAT RECH SCIENT (FR); PASTEUR INSTITUT (FR)

Requested Patent:  EP0750041, A3

Application Number: EP19960108720 19880115

Priority Number (s): EP19880400084 19880115; FR19870001739 19870211; FR19870005398 19870415; US19870003764 19870116

IPC Classification: C12N15/49 ; C07K14/16 ; C12Q1/68

EC Classification: C07K14/16B, C07K14/16D, C07K14/16, C12Q1/70B2B

Equivalents:

### Abstract

New peptides (I) have immunological properties in common with those of the peptide skeleton of the envelope protein of HIV-2 and also have a peptide structure in common with that of SIV (simian immunodeficiency virus)-7 glycoprotein. Also new are (i) a sequence of 9600 nucleotides corresponding to the SIV genome (reproduced in the specification, together with derived amino acid sequence of viral proteins gag, pol, env, Q, X, R, tat, art and F genes) and its fragments; (2) recombinant DNA contg. all or part of the CDNA from this sequence inserted into a vector, and (3) antigenic and immunogenic conjugates contg. (I).

Data supplied from the esp@cenet database - I2



(19)

Europäisches Patentamt  
European Patent Office  
Office européen des brevets



(11)

EP 0 750 041 A2

(12)

## DEMANDE DE BREVET EUROPEEN

(43) Date de publication:  
27.12.1996 Bulletin 1996/52

(51) Int Cl.6: C12N 15/49, C07K 14/16,  
C12Q 1/68

(21) Numéro de dépôt: 96108720.2

(22) Date de dépôt: 15.01.1988

(84) Etats contractants désignés:  
AT BE CH DE ES FR GB GR IT LI LU NL SE

(30) Priorité: 16.01.1987 US 3764  
11.02.1987 FR 8701739  
15.04.1987 FR 8705398

(62) Numéro de dépôt de la demande initiale en  
application de l'article 76 CBE: 88400084.5

(71) Demandeurs:  
• INSTITUT PASTEUR  
F-75015 Paris (FR)  
Etats contractants désignés:  
BE CH DE ES GB GR IT LI LU NL SE AT  
• CENTRE NATIONAL DE  
LA RECHERCHE SCIENTIFIQUE  
75007 Paris (FR)  
Etats contractants désignés:  
FR

(72) Inventeurs:  
• Alizon, Marc  
75005 Paris (FR)  
• Montagnier, Luc  
92350 Le Plessis Robinson (FR)

- Guetard, Denise  
75015 Paris (FR)
- Clavel, Françoise  
Rockville, MD 20852 (US)
- Sonigo, Pierre  
75015 Paris (FR)
- Guyader, Mireille  
75017 Paris (FR)
- Tlollals, Pierre  
75013 Paris (FR)
- Chakrabarti, Lisa  
750 Paris (FR)
- Desrosiers, Ronald  
Hudson, Massachusetts 01749 (US)

(74) Mandataire: Desalix, Anne  
Ernest Gutmann - Yves Plasseraud S.A.  
3, rue Chauveau-Lagarde  
75008 Paris (FR)

## Remarques:

Cette demande a été déposée le 31 - 05 - 1996  
comme demande divisionnaire de la demande  
mentionnée sous le code INID 62.

## (54) Séquences de nucléotides de HIV-2

(57) L'invention concerne une séquence de nucléotides caractérisée en ce qu'elle répond à la séquence nucléotidique représentée à la figure 1B ou à la figure 1C, ou en ce qu'elle contient la séquence nucléotidique représentée à la figure 1B ou à la figure 1C, ou en ce qu'il s'agit d'une partie de la séquence représentée à la

figure 1B ou à la figure 1C, ladite partie de séquence codant pour un peptide reconnu par des anticorps présents dans le sérum d'un patient infecté par un rétrovirus HIV-2 ou étant utilisable comme sonde pour la détection dans un échantillon biologique, de la présence d'un rétrovirus HIV-2.

EP 0 750 041 A2

**Description**

La présente invention est relative à des peptides ayant des propriétés immunologiques, le cas échéant immuno-gènes, en commun avec des antigènes susceptibles d'être obtenus sous une forme purifiée, à partir de virus capables de provoquer des lymphadénopathies susceptibles d'être relayées ensuite par le syndrome d'immunodéficiency acquise (SIDA) chez l'homme.

L'invention concerne en particulier des peptides antigéniques susceptibles d'être reconnus par des anticorps induits chez l'homme par des virus désignés par l'abréviation HIV, selon la nomenclature définie dans NATURE. Elle concerne également des peptides ayant des propriétés immunogènes ou susceptibles d'être rendus immunogènes *in vivo*, cette immunogénicité étant susceptible de se manifester par l'induction *in vivo* d'anticorps reconnaissant des antigènes caractéristiques des virus HIV-2 et même, au moins en ce qui concerne certains de ces peptides, des antigènes issus de HIV-1.

L'invention concerne en outre des applications de ces peptides à la fabrication de compositions pour le diagnostic *in vitro* chez l'homme de potentialité de certaines formes du SIDA et, en ce qui concerne certains d'entre eux, à la production de compositions immunogènes et de compositions vaccinantes contre les rétrovirus HIV.

De même l'invention concerne les applications aux mêmes fins des anticorps susceptibles d'être induits *in vivo* par les peptides immunogènes ou rendus immunogènes et, pour certains de ces anticorps, leurs applications à la production de principes actifs de médicaments contre ces SIDAS humains.

L'invention concerne également la mise en œuvre de certains de ces peptides dans des procédés pour le diagnostic *in vitro* chez l'homme de certaines formes du SIDA, ainsi que leur application à la constitution de trousseaux ou "kits" de diagnostic.

Un premier rétrovirus dénommé LAV-1 ou HIV-1 a été isolé et décrit dans la demande de brevet GB.83/24.800 et une demande EP.84/401.834 du 14/09/84. Ce virus a également été décrit par F.Barre Sinoussi et al. dans Science, 220 n° 45-99, 20 pages 868-871.

Des variants de ce virus HIV-1 désignés par LAV ELI et LAV MAL, ont également été isolés, caractérisés et décrits dans la demande de brevet EP.84-401.834.

Les virus HIV-1 et leurs variants possèdent les propriétés suivantes :

- ils ont pour cibles préférentielles les cellules Leu3 (ou lymphocytes T4) humaines et leurs cellules dérivées "immortalisées".
- ils ont une activité transcriptase inverse nécessitant la présence d'ions Mg<sup>2+</sup> et présentent une forte activité pour le poly(adenylate-oligo-deoxythymidylase) poly(A)-oligo(dT)12-18)
- ils ont une densité de 1,16 à 1,17 sur gradient de sucre,
- ils ont un diamètre moyen de 139 nanomètres et un noyau de diamètre moyen de 41 nanomètres,
- les lysats de ces virus contiennent une protéine p25 (protéine du noyau) qui ne croise pas immunologiquement avec la protéine p24 de HTLV-1,
- ils contiennent une protéine p42 appartenant à leur enveloppe,
- ils contiennent également une glycoprotéine d'enveloppe gp110 d'un poids moléculaire de 110.000.

L'isolement et la caractérisation de rétrovirus appartenant à une classe distincte et n'ayant qu'une parenté immunologique réduite avec les précédents, ont été décrits dans la demande de brevet européen n° 87/400.151.4. Ces rétrovirus qui ont été regroupés sous la désignation HIV-2, ont été isolés chez plusieurs malades africains présentant des symptômes d'une lymphadénopathie ou d'un SIDA.

Les rétrovirus du type HIV-2 comme les rétrovirus du type HIV-1, se caractérisent par un tropisme pour les lymphocytes T4 humains et par un effet cytopathogène à l'égard de ces lymphocytes, lorsqu'ils s'y multiplient, pour alors causer soit des poly-adénopathies généralisées et persistantes, soit un SIDA.

Plus généralement les rétrovirus purifiés par HIV-2 possèdent en général les propriétés suivantes :

- la cible préférentielle des rétrovirus HIV-2 est constituée par les cellules Leu3 (ou lymphocytes T4) humaines et pour des cellules "immortalisées" dérivées de ces lymphocytes T4 ;
- ils sont cytotoxiques pour les lymphocytes T4 humains
- ils ont une activité de transcriptase inverse nécessitant la présence d'ions Mg<sup>2+</sup> et présentant une forte activité pour le poly(adénylate-oligodéoxythymidylase) (poly(A)-oligo(dT) 12-18) ;
- ils ont une densité de 1,16 dans un gradient de sucre ;
- ils ont un diamètre moyen de 140 nanomètres et un noyau ayant un diamètre moyen de 41 nanomètres ;
- ils peuvent être cultivés dans des lignées permanentes du type HUT ou exprimant la protéine T4 ;
- ils ne sont pas infectieux pour les lymphocytes T8 ;
- les lysats de ces virus contiennent une protéine p26 qui ne croise pas immunologiquement avec la protéine p24

- du virus HTLV-I ou du virus HTLV-II ;
- ces lysats contiennent en outre une protéine p16 qui n'est pas reconnue immunologiquement par la protéine p19 de HTLV-I ou de HTLV-II dans des essais de radioimmuno-précipitation ;
- ils contiennent en outre une glycoprotéine d'enveloppe ayant un poids moléculaire de l'ordre de 130.000-140.000 qui ne croise pas immunologiquement avec la gp110 des HIV-1, mais qui en revanche croise immunologiquement avec la glycoprotéine d'enveloppe gp140 de STLV-III (virus isolé chez le singe) ;
- ces lysats contiennent encore des antigènes marquables par la <sup>35</sup>S-cystéine, dont les poids moléculaires s'étagent entre 32.000 et 42.000-45.000 : ils comprennent notamment un antigène ayant un poids moléculaire de l'ordre de 36.000 et un antigène ayant un poids moléculaire de l'ordre de 42.000, l'un de ces antigènes (p36 et p42) constituant vraisemblablement une glycoprotéine transmembranaire du virus HIV-2 ;
- l'ARN génomique des HIV-2 n'hybride pas avec l'ARN génomique de HIV-1 dans des conditions stringentes ;
- dans des conditions non stringentes, l'ARN génomique de HIV-2 n'hybride, ni avec le gène env et le LTR qui le jouxte, de HIV-1, ni avec des séquences de la région pol du génome de HIV-1 ;
- dans des conditions non stringentes, il hybride faiblement avec des séquences de nucléotides de la région de HIV-1.

Un autre rétrovirus dénommé SIV-1, cette dénomination remplaçant la dénomination antérieurement connue STLV III, a été isolé chez le singe macaque rhésus. (M.D.Daniel et al. Science 228, 1201 (1985) N.L.Letwin et al, Science 230, 71 (1985) sous l'appellation "STLV-III<sub>mac</sub>").

Un autre rétrovirus, désigné "STLV-III<sub>AGM</sub>", (ou SIV<sub>AGM</sub>) a été isolé chez des singes verts sauvages. Mais, contrairement au virus présent chez le singe macaque rhésus, la présence de "STLV-III AGM" ne semble pas induire une maladie du type SIDA chez le singe vert d'Afrique.

Une souche du rétrovirus SIV-1mac a été déposée à la CNCM le 7 Février 1986 sous le n° I-521. Des études ont montré que le rétrovirus SIV-1 comporte certaines protéines possédant une certaine parenté immunologique avec des protéines ou glycoprotéines structurales susceptibles d'être obtenues dans des conditions analogues, à partir de HIV-2. Ce rétrovirus SIV-1, dont on a constaté le caractère infectieux chez les singes, avait été désigné par STLVIII par les chercheurs qui l'ont isolé (références bibliographiques précitées).

Pour la commodité du langage, ces virus ne seront plus désignés dans ce qui suit que par l'expression SIV (l'expression SIV est l'abréviation anglaise de "Simian Immunodeficiency Virus" (virus d'immunodéficience du singe) éventuellement suivi d'une abréviation désignant l'espèce de singe dont ils sont issus, par exemple, MAC (ou mac) pour le macaque ou AGM pour le singe vert d'Afrique (abréviation de "African Green Monkey").

En mettant en oeuvre les mêmes techniques que celles rappelées plus haut, il a été constaté que l'on pouvait également obtenir à partir de SIV-1mac :

- une protéine principale du noyau p27, ayant un poids moléculaire de l'ordre de 27 kilodaltons,
- une glycoprotéine majeure d'enveloppe, gp140,
- une protéine vraisemblablement transmembranaire p32, qui n'est guère observée en RIPA lorsque le virus a au préalable été marqué par la <sup>35</sup>S-cystéine, mais qui peut être observée dans les essais d'immunoempreintes (Western blots), sous forme de bandes larges.

Des études plus précises ont été réalisées en ce qui concerne les précédents virus HIV-2 et SIV. La poursuite de l'étude des rétrovirus HIV-2 a également conduit à l'obtention de séquences d'ADN complémentaires (ADNc) des ARNs de leurs génomes. La séquence nucléotidique complète de l'ADNc d'un rétrovirus représentatif de la classe HIV-2 (HIV-2 ROD) a été déposée le 21/02/1986 à la CNCM sous le n° I-522, sous le nom de référence LAV-II ROD.

Cette séquence nucléotidique et les phases de lecture ouverte qu'elle contient sont indiqués à la figure 1 A.

En outre, la poursuite de l'étude d'autres rétrovirus a également permis d'aboutir à l'obtention de leurs séquences nucléotidiques complètes. Il en est en particulier ainsi de l'ADNc dérivé de l'ARN génomique de SIV.

Le clonage et le séquençage du virus SIV-1mac qui ont permis l'obtention de sa séquence nucléotidique ont été réalisés dans les conditions suivantes :

L'ADN de cellules HUT 78 infectées par le virus SIV (isolat STLV-III mac 142-83 décrit par Daniel et al.(1985) Science, 228, p.1201-1204, digéré partiellement par l'enzyme de restriction Sau3A a été cloné au site BamHI du bactériophage vecteur Lambda ELBL3 pour constituer une banque génomique. Les 2 millions de phages recombinants de la banque génomique ainsi constituée ont été criblés *in situ* en conditions de sécurité P3, à l'aide de séquences du virus HIV2 provenant des clones lambda-ROD4, lambda-ROD35 et E2 (Clavel et al. (1986-Nature, 324, p.691.) et nick-translatées.

L'hybridation a été réalisée en 5xSSC à 50°C et les lavages en 2xSSC à 5°C. Un seul clone contenant l'ensemble des séquences virales a été obtenu. Ce clone est désigné par lambda-SIV-1. L'insérat du phage lambda-SIV-1 mesure 16,5 kb au total et comprend un provirus intégré auquel manquent seulement les 250 premières bases du LTR gauche,

alors que le LTR droit est complet.

Le provirus intégré a été séquencé par la méthode des didéoxynucléotides après sous-clonage de fragments aléatoires dans le phage M13mp8. 300 sous-clones ont été analysés.

Des fragments d'ADNc provenant du clone Lambda SIV-1 insérés dans des plasmides pSIV-1.1 et pSIV-1.2 ont été déposés à la CNCM le 15 Avril 1987, sous les numéros I-658 (pSIV-1.1) et I-659 (pSIV-1.2).

Les résultats ont été mentionnés dans les figures décrites ci-après.

La figure 1B représente la séquence nucléotidique du génome viral de SIV et les séquences qui en sont déduites pour les protéines virales correspondant aux produits des gènes gag, pol, env, Q, X, R, tat, art, F.

Les figures 3 à 11 et la figure 1C représentent les comparaisons des produits théoriques des gènes viraux et des LTR entre HIV2 et SIVmac. (λSIV-1).

L'invention concerne de plus les fragments d'ADNc déduits de l'ADNc issu du génome entier de SIV-1, ces fragments contenant une ou plusieurs séquences issues de la séquence complète d'ADNc et qui codent pour des peptides intéressants de l'invention. Ces séquences sont indiquées à la figure 1B et, à la figure 1C pour ce qui a trait à la séquence LTR du virus.

Les séquences nucléotiques de l'ADNc de SIV ont été placées en correspondance avec les séquences nucléotiques du virus HIV-2 ROD pour ce qui concerne la séquence LTR (figure 1C). Cette présentation que l'on retrouve pour le génome entier en rapprochant la figure 1B des figures 3 à 11 permet de repérer ou de déduire les acides nucléotiques ayant des éléments de structure essentiels communs aux deux virus.

L'invention concerne naturellement aussi l'utilisation des cADNs issus de SIV ou de leurs fragments (ou de recombinants les contenant) en tant que sondes, pour le diagnostic de la présence ou non de virus HIV-2 dans des échantillons de sérums ou d'autres liquides ou tissus biologiques obtenus à partir de patients suspectés d'être porteurs du virus HIV-2. Ces sondes sont de préférence marquées également (marqueurs radio-actifs, enzymatiques, fluorescents, etc.). Des sondes particulièrement intéressantes pour la mise en œuvre du procédé de diagnostic du virus HIV-2 ou d'un variant de HIV-2 peuvent être caractérisées en ce qu'elles comprennent la totalité ou une fraction de l'ADNc complémentaire du génome du virus SIV ou encore notamment les fragments recombinants contenus dans divers clones.

Les sondes mises en œuvre dans ce procédé de diagnostic du virus HIV-2 et dans les kits de diagnostic ne sont en aucune façon réduites aux sondes décrites précédemment. Elles comprennent au contraire toutes les séquences nucléotidiques issues du génome du virus SIV, d'un variant de SIV ou d'un virus proche par sa structure, dès lors qu'elles permettent la détection dans des fluides biologiques de personnes susceptibles de développer un SIDA, d'anticorps dirigés contre un HIV-2 ou d'un virus qui en est proche.

La détection peut être réalisée de toutes façons en soi connues. Elle peut comprendre une mise en contact de ces sondes soit avec les acides nucléotiques obtenus à partir des cellules contenues dans ces sérums ou autres milieux biologiques, par exemple liquides céphalo-rachidiens, salives, etc... Elle peut aussi comprendre une mise en contact de ces sondes avec ces milieux eux-mêmes dès lors que leurs acides nucléotiques ont été rendus accessibles à l'hybridation avec ces sondes, et ce dans des conditions permettant l'hybridation entre ces sondes et ces acides nucléotiques. L'étape finale du diagnostic *in vitro* comprend alors la détection de l'hybridation éventuellement produite. Le susdit diagnostic mettant en jeu des réactions d'hybridation peut également être réalisé à l'aide de mélanges de sondes respectivement originaires d'un HIV-2 et d'un SIV-1 ou d'un HIV-1, d'un HIV-2 et d'un SIV, dès lors qu'il n'est pas nécessaire de faire une différence entre le type de virus recherché.

D'une façon générale, le procédé de diagnostic de la présence ou non du virus HIV-2 ou d'un variant dans des échantillons de sérums ou d'autres liquides ou tissus obtenus à partir de patients suspectés d'être porteurs du virus HIV-2 comprend les étapes suivantes :

- 45 1/ au moins une étape d'hybridation conduite dans des conditions stringentes, par mise en contact de l'ADN de cellules de l'échantillon du patient suspect avec l'une des susdites sondes marquées sur une membrane appropriée,
- 46 2/ le lavage de ladite membrane avec une solution assurant la conservation de ces conditions stringentes de l'hybridation,
- 50 3/ la détection de la présence ou non du virus HIV-2 par une méthode d'immunodétection.

Dans un autre mode de réalisation préféré du procédé selon l'invention l'hybridation précitée est conduite dans des conditions non stringentes et le lavage de la membrane est réalisé dans des conditions adaptées à celles de l'hybridation.

Il va de soi que l'invention concerne les acides nucléotiques correspondant à des séquences placées en des régions analogues de variants de SIV ainsi que tous les acides nucléotiques dont les modifications résulteraient de la mise à profit de la dégénérescence du code génétique.

Les études comparatives qui ont aussi permis d'aboutir à des résultats relatifs aux protéines de noyau (core), ci-

après dénommées "protéines gag" et aux protéines d'enveloppes, ci-après dénommées "protéines env", ont également été rapportés dans la demande de brevet européen n° 87/400.151.4, déjà citée. Ces résultats montrent que les protéines du noyau (protéines gag) dans HIV-2 présentent des différences moins accentuées par rapport à celles des virus HIV-1, que les protéines d'enveloppe (protéines env). Globalement les protéines env dans HIV-2 se sont révélées présenter des parentés immunologiques extrêmement faibles, sinon inexistantes, avec les protéines env correspondantes des virus HIV-1.

5 Au contraire des études comparatives effectuées entre les structures des séquences d'ADNc des virus HIV-2 et SIV permettent de mettre en évidence certaines caractéristiques communes qui apparaissent au niveau des protéines. Globalement, les protéines de HIV-2 et de SIV-1 montrent des parentés immunologiques importantes.

10 La glycoprotéine majeure d'enveloppe de HIV-2 s'est révélée être plus proche immunologiquement de la glycoprotéine majeure d'enveloppe de SIV que de la glycoprotéine majeure d'enveloppe de HIV-1.

15 Ces constatations s'imposent non seulement au niveau des poids moléculaires : 130-140 kilodaltons pour les glycoprotéines majeures de HIV-2 et de SIV contre environ 110 pour la glycoprotéine majeure d'enveloppe de HIV-1, mais aussi au niveau des propriétés immunologiques, puisque des sérum prélevés à partir de malades infectés par HIV-2, et plus particulièrement des anticorps formés contre la gp140 de HIV-2 reconnaissent la gp140 de SIV-1mac, alors que dans des essais semblables les mêmes sérum et les mêmes anticorps de HIV-2 ne reconnaissent pas la gp110 de HIV-1. Mais les sérum anti-HIV-1 qui n'ont jamais réagi avec la gp140 de HIV-2 précipitent une protéine de 26 Kdal marquée par la <sup>35</sup>S-cystéine, contenue dans les extraits de HIV-2.

20 La protéine majeure du noyau (core) de HIV-2 semble présenter un poids moléculaire moyen (environ 26.000) intermédiaire entre celui de la p25 de HIV-1 et la p27 de SIV.

Ces observations résultent des essais réalisés avec des extraits viraux obtenus à partir du HIV-2 isolé à partir de l'un des patients susmentionnés. Des résultats similaires ont été obtenus avec des extraits viraux du HIV-2 isolé à partir du second patient.

25 Des études plus poussées ont conduit les inventeurs à reconnaître une première classe de peptides ayant des séquences d'aminoacides soit identiques, soit proches de séquences contenues à l'intérieur des structures des protéines gag et env de HIV-2 ou de SIV voire de HIV-1. Ces peptides sont notamment applicables au diagnostic d'une infection chez l'homme par le virus HIV-2 ou de l'un de ses variants.

30 A cet égard la présente invention concerne également des procédés et des compositions de diagnostic pour la détection in vitro d'anticorps dirigés contre un virus HIV-2 ou de ses variants, plus particulièrement dans des échantillons biologiques, notamment des sérum de patients ayant subi une infection par le virus HIV-2, certains de ces peptides permettant une discrimination particulièrement poussée entre les infections dues à des virus HIV-2 et à des virus HIV-1.

35 Ces études poussées ont également conduit à la possibilité de synthétiser des peptides immunogènes ou susceptibles d'être rendus immunogènes, présentant des caractéristiques de structures leur permettant d'induire in vivo la production d'anticorps susceptibles de reconnaître des protéines env à la fois dans HIV-1 et dans HIV-2 et, au moins pour certains de ces peptides, de se fixer tant sur des virus HIV-1 que sur des virus HIV-2, plus particulièrement aux fins de les neutraliser. L'utilisation de ces derniers types de peptides est donc particulièrement indiquée pour la production de principes actifs de vaccins contre les virus HIV, donc contre le SIDA.

40 Pour désigner ci-après les résidus d'aminoacides entrant dans la constitution des peptides selon l'invention, on aura recours, pour ceux des acides aminés ayant une signification unique à la nomenclature internationale désignant chaque acide aminé naturel par une lettre unique (lettre majuscule) selon le tableau des correspondances qui suit :

M Méthionine

L Leucine

I Isoleucine

45 V Valine

F Phenylalanine

S Sérine

P Proline

T Thréonine

50 A Alanine

Y Tyrosine

H Histidine

Q Glutamine

N Asparagine

55 K Lysine

D Acide Aspartique

E Acide glutaminique

C Cystéine

**EP 0 750 041 A2**

W Tryptophane  
R Arginine  
G Glycine

- 5 Lorsqu'un acide aminé pourra, en raison de sa position au sein de la chaîne d'aminoacides caractéristique d'un peptide déterminé, prendre plusieurs significations, il pourra soit être désigné par un tiret "-", si sa signification peut être quelconque, soit par une lettre minuscule lorsque cet aminoacide pourra présenter un nombre limité de significations préférées, ce nombre étant cependant toujours supérieur à 1. Dans ce dernier cas, les significations possibles de cette lettre minuscule seront toujours précisées en rapport avec le peptide auquel il appartient.
- 10 Afin de faciliter la lecture, ces peptides seront désignés par une abréviation env ou gag suivie d'un indice numérique, par référence à des séquences d'aminoacides contenues, selon le cas, soit dans les protéines env soit dans les protéines gag de certains HIV-1, HIV-2 ou SIV. Il y sera encore fait référence dans ce qui suit.  
Enfin dans les définitions qui suivent
- 15 - les groupes X représentent soit un groupe NH<sub>2</sub> libre ou amidé, notamment par un ou deux groupes alcoyle comprenant de 1 à 5 atomes de carbone, soit un groupe peptidique comprenant de 1 à 5 aminoacides, dont l'aminoacide N-terminal présente lui-même un groupe NH<sub>2</sub> libre ou amidé comme précédemment indiqué, et
- les groupes Z représentent, soit un groupe -OH libre ou alcoxyle et contenant alors un groupe alcoyle comprenant de 1 à 5 atomes de carbone, soit un groupe peptidique comprenant de 1 à 5 aminoacides, dont l'aminoacide C-terminal présente lui-même un groupe -OH libre ou alcoxyle, comme précédemment indiqué, les groupes de 1 à 5 acides aminés le cas échéant contenus dans X ou Z ou dans les deux à la fois étant tels, que leur présence n'est pas incompatible avec la préservation pour l'essentiel des propriétés immunologiques, le cas échéant immunogènes, des peptides qui en sont dépourvus.
- 20
- 25 Les peptides selon l'invention, qui ont en commun des propriétés immunologiques avec des antigènes de HIV-2 et, pour certains d'entre eux également avec des antigènes de HIV-1 ou de ses variants, sont caractérisés en ce qu'ils ont également une structure peptidique en commun avec les antigènes de SIV. De façon avantageuse, ces peptides comprennent normalement au plus 40 résidus d'acides aminés.
- Des peptides préférés sont les suivants :

30

35

40

45

50

55

5           env1  
XRV-AIEKYL-DQA-LN-WGCAFRQVCZ  
10          env2  
X-LE-AQI-QQEKNMYELQKLNZ  
15          env3  
XELGDYKLVEITPIG-APT--KR----Z  
20          env4  
X----VTV-YGVP-WK-AT--LFCA-Z  
25          env5  
X---QE--L-NVTE-F--W-NZ  
30          env6  
XL---S-KPCVKLTPLCV--Z  
35          env7  
X---N-S-IT--C-K----Z  
40          env8  
X-I---YC-P-G-A-L-C-N-TZ  
45          env9  
X-----A-C-----W--Z  
50          env10  
X-G-DPE-----NC-GEF-YCN----NZ

35          env11  
X-----C-IKQ-I-----G---YZ

40          Plus particulièrement l'invention concerne les peptides suivants :

45

50

55

5           env1  
XRV-AIEKYL-DQA-LN-WGCAFRQVCZ  
10          env2  
X-LE-AQIQQEKNMYELQKLNZ  
15          env3  
XELGDYKLVEITPIG-APT--KR----Z  
20          env4  
X----VTV-YGVP-W--AT--LFCA-Z  
25          env5  
X----E--L-NVTE-F--W-NZ  
30          env6  
XL---S-KPCVKL-PLC---Z  
35          env7  
X---N-S-I---C-K----Z  
40          env8  
X-I---YC-P-G-A-L-C-N-TZ  
45          env9  
X-----A-C-----W--Z  
50          env10  
X-G-DPE-----NC-GEF-YC-----NZ  
55          env11  
X-----C-I-Q-I-----G---YZ

Des peptides avantageux correspondant aux précédents, présentent les formules qui suivent :

50          env1  
XRVTAIEKYLQDQARLN SWGCAFRQVCZ, ou  
XRVTAIEKYLKDQQLNAWGCAFRQVCZ  
55          env2  
XSLEQAQIQQEKNMYELQKLN SWZ, ou  
KLLEEAQIQQEKNMYELQKLN SWZ  
60          env3  
XELGDYKLVEITPIGFAPTKEKRYSSAHZ, ou  
XELGDYKLVEITPIGLAPTNVKRYTTG-Z

65          (On remarquera que les peptides env1, env2, env3 attestent de la très grande parenté entre HIV-2 et SIV-1. En effet le premier peptide est inclus dans le génome de HIV-2 et le second, dans celui de SIV-1).

env4

XabcdVTVeYGV PfWogAThiLFCAjZ,

5 dans lesquels les lettres de a à j peuvent avoir les significations suivantes :

- a est C, E ou D
- b est T, K, D, N ou I
- c est Q ou L
- 10 d est Y ou W
- e est F ou Y
- f est T, V ou A
- g est N ou E
- h est I ou T
- i est P ou T
- 15 j est T ou S
- o est K ou R

20 env5

XabcoEdeLfNVTEgFhiWjNZ,

25 dans lequel les lettres de a à j peuvent avoir les significations suivantes :

- a est D ou P
- b est D ou N
- c est Y ou P
- d est I, V, I ou L
- 30 e est T, V, E ou A
- f est V, G ou E ou -
- g est A, N, G ou S
- h est D ou N
- i est A ou M
- j est N, K ou E
- 35 o est Q ou S

env6

XLabcSdKPCVKLoPLCuefKZ,

40

dans lequel les lettres de a à f peuvent avoir les significations suivantes :

- 45 a est F ou W
- b est E ou D
- c est T ou Q
- d est I ou L
- e est A, S ou T
- f est M ou L
- 50 o est T ou S
- u est V ou I

env7

XabCNxSyIocdCeKfghiz,

55

dans lequel les lettres de a à i et x et y peuvent avoir les significations suivantes :  
a est N ou T ou I

**EP 0 750 041 A2**

b est H ou S ou N  
c est E ou Q

d est S, A ou C  
e est D ou P

5 f est H, V ou D

g est Y ou S

h est W ou F

i est D ou E

x est T ou R

10 y est V ou A

o est T ou Q

**env8**

15 **XaIbcdYCxPeGfAgLhCiNjtZ,**

dans lequel les lettres de a à k et x peuvent avoir les significations suivantes :

a est A ou P

20 b est R ou P

c est F, I ou C

d est R ou H

e est P ou A

f est Y ou F

25 g est L ou I

h est R ou K

i est - ou N

j est D ou K

x est A ou T

30

**env9**

**XwabcxxyAdCefghizWjkz,**

35

dans lequel les lettres de a à k et x à z peuvent avoir les significations suivantes :

a est K ou - ou E

b est R ou -

c est P ou M ou I

40 d est W ou H ou Y

e est W ou N ou T ou R

f est F ou I

g est K ou S ou N ou G

h est G ou R ou E

45 i est - ou A ou T

j est K ou N ou D ou S

k est D ou A ou N ou K ou E

w est N, D ou I

x est R ou G ou K

y est Q ou K ou R

50 z est K ou E ou Q ou N

**env10**

**XaGbDPEcdefghNCiGEFjYCokxlmnNZ,**

dans lequel les lettres de a à n et x peuvent avoir les significations suivantes :

a est K ou - ou G  
b est S ou G ou -  
c est V ou I  
d est A ou V ou T  
5 e est Y ou T ou M ou F  
f est M ou H  
g est W ou S  
h est T ou F  
i est R ou G  
j est L ou F  
10 o est N ou K  
k est M ou S  
l est W ou Q ou K ou G  
m est F ou L  
n est L ou F  
x est T ou S ou N

20 env11  
**XabcdwCeIoQfIxgyhizGjklyZ,**

dans lequel les lettres de a à l et w à z peuvent avoir les significations suivantes :  
a est R ou T ou S ou N  
25 b est N ou I  
c est Y ou T  
d est A ou L ou V  
e est H ou R  
f est I ou F  
30 g est T ou M  
h est H ou Q ou A  
i est K ou E  
j est R ou K  
k est N ou A  
35 l est V ou M  
w est P ou Q  
x est N ou K  
y est W ou V  
z est V ou T ou K  
40 o est K ou R

La structure du peptide antigénique codé par le gène gag et désigné par gag1 est également représentée ci-après :

45 **XDCKLVLKGLGaNPTLEEMLTAZ ,**

dans lequel la lettre a désigne M ou T.  
Il sera remarqué que, d'une façon générale, les aminoacides ayant une signification univoque (donc représentés par une lettre majuscule correspondant à la nomenclature internationale) qui interviennent dans les définitions qui précèdent des peptides selon l'invention, se trouvent être la correspondance avec des aminoacides identiques placés dans le même ordre dans les séquences env ou gag correspondantes de la protéine env ou gag d'au moins l'un des HIV, ou de SIV-1.

Les positions de ces séquences sont soulignées et repérées au sein des séquences d'aminoacides des protéines env respectivement de HIV-2 ROD (CNCM n° I-532) et HIV-1 BRU (CNCM n° I-232) représentées à la figure 2. Par ailleurs, les alignements des acides aminés des protéines env et gag respectivement de SIV-1mac (CNCM n° I.521) et de HIV-2 ROD sont présentées à la figure 3 et à la figure 4.

Les traits pleins qui apparaissent en certaines localisations de ces séquences visent à souligner que certains aminoacides contenus dans ces séquences ont été volontairement déletés au plan de la présentation, afin de permettre

EP 0 750 041 A2

la mise en alignement d'aminoacides respectivement identiques (alors marqués d'un astérisque) ou de deux points verticaux sur une même ligne verticale dans les séquences des protéines correspondantes de HIV-1 et de HIV-2 d'une part, de SIV et de HIV-2 d'autre part.

Outre les peptides précités, l'invention concerne également les peptides modifiés par insertion et/ou délétion et/ou substitution d'un ou plusieurs acides aminés, pour autant que les propriétés antigéniques ou immunogènes desdits peptides ne sont pas modifiées, ou que les propriétés de reconnaissance de l'antigène ou de l'anticorps avec lesdits peptides ne sont pas substantiellement modifiées.

Dans un mode de réalisation particulièrement préféré, l'invention concerne des peptides ayant des propriétés immunologiques en commun avec l'ossature peptidique de la glycoprotéine d'enveloppe des virus de la classe HIV-2, ces peptides contenant un nombre de résidus d'acides aminés n'excédant pas 40.

Ces peptides préférés selon l'invention ont les séquences suivantes :

env1

15           RVTAIEKYLQDQARLNSWGCAFQVC  
              AIEKYLQDQ  
              RVSAIEKYLKDQAQLNAWGCAFQVC  
              AIEKYLKDQ

20

env2

25           SLEQAQIQQEKNMYELQKLNSW  
              QIQQEKN  
              LLEEAQIQQEKNMYELQKLNSW

30           env3

              ELGDYKLVEITPIGFAPTKEKRYSSAH  
              YKLVEITPIGFAPTKEK  
              ELGDYKLVEITPIGLAPTNVKRYTTG-  
              YKLVEITPIGLAPTNVK

35

40           env4  
              CTQYVTVFYGVPTWKATIPLFCAT  
              VTVFYGVPTWKAT  
              CIQYVTVFYGVPAWRNATIPLFCAT  
              VTVFYGVPAWRNAT

45

50           EKLWVTVYYGVPVWKEATTLFCAS  
              VTVYYGVPVWKEAT  
              EDLWVTVYYGVPVWKEATTLFCAS  
              VTVYYGVPVWKEAT  
              DNLWVTVYYGVPVWKEATTLFCAS  
              VTVYYGVPVWKEAT

55

5                   env5  
DDYQEITL-NVTEAFDAWN  
L-NVTEAF  
DDYSELAL-NVTESFDAWEN  
L-NVTESF  
10                 PNPQEVVLVNVTENFNMWKN  
LVNVTENF  
PNPQEIELENVTATEGFNMWKN  
LENVTEGF  
15                 PNPQEIALENVTENFNMWKN  
LENVTENF  
  
20                 env6  
ETSIKPCVKLTPLCVAMK  
ETSIKPCVKLSPLCITMR  
25                 DQSLKPCVKLTPLCVSLK  
DQSLKPCVKLTPLCVTLN  
PCVKLTPLCV  
  
30                 env7  
NHCNTSVITESCD  
NTSVIT  
35                 NHCNTSVIQECCD  
NTSVIQ  
TSCNTSVITQACP  
NTSVIT  
40                 INCNTSVITQACP  
NTSVIT  
INCNTSAITQACP  
NTSAIT  
  
45  
  
50                 env8  
YCAPPGYALLRC-NDT  
YCAPAGFAILKCNNKT  
55                 YCAPAGFAILKCNDKK  
YCAPAGFAILKCRDKK

env9

5 NKRPQAWCWFKG-KWKD  
NERPKQAWCRFGG-NWKE  
N--MRQAHCNISRAKWNA  
D--IRRAYCTINETEWDK  
10 I--IGQAHCNISRAQWSK

env10

15 KGSDPEVAYMWTNCRGEFLYCNMTWFLN  
NCRGEFLYCN  
GG-DPEVTMWTNCRGEFLYCKMNWFLN  
NCRGEFLYCK  
20 -GGDPEIVTHSFNCGGEFFYCNSTQLFN  
NCGGEFFYCN  
-GGDPEITTHSFNCRGGEFFYCNTSKLFN  
NCRGEFFYCN  
25 -GGDPEITTHSFNCGGEFFYCNTSGLFN  
NCGGEFFYCN

env11

30 RNYAPCHIKQIINTWHKVGRNVY  
CHIKQII  
35 RNYVPCHIRQIINTWHKVGNVY  
CHIRQII  
TITLPCRIKQFINMWQEVGKAMY  
40 CRIKQFI  
SITLPCRIKQIINMWQKTCKAMY  
CRIKQII  
45 NITLQCRIKQIIKMVAGR-KAIY  
CRIKQII

gag1

50 DCKLVLKGLGTNPTELMLTA

Les peptides selon l'invention peuvent encore avantageusement être préparés par les techniques classiques, dans le domaine de la synthèse des peptides. Cette synthèse peut être réalisée en solution homogène ou en phase solide.

55 Par exemple, on aura recours à la technique de synthèse en solution homogène décrit par HOUBENWEYL dans l'ouvrage intitulé "Méthode der Organischen Chemie" (Méthode de la Chimie Organique) édité par E. Wunsch, vol. 15-I et II., THIEME, Stuttgart 1974.

Cette méthode de synthèse consiste à condenser successivement deux-à-deux les aminoacyles successifs dans

- l'ordre requis, ou à condenser des aminoacyles et des fragments préalablement formés et contenant déjà plusieurs aminoacyles dans l'ordre approprié, ou encore plusieurs fragments préalablement ainsi préparés, étant entendu que l'on aura eu soin de protéger au préalable toutes les fonctions réactives portées par ces aminoacyles ou fragments, à l'exception des fonctions amines de l'un et carboxyles de l'autre ou vice-versa, qui doivent normalement intervenir
- 5 dans la formation des liaisons peptidiques, notamment après activation de la fonction carboxyle, selon les méthodes bien connues dans la synthèse des peptides. En variante, on pourra avoir recours à des réactions de couplage mettant en jeu des réactifs de couplage classique, du type carbodiimide, tels que par exemple la 1-éthyl-3-(3-diméthyl-amino-propyl)-carbodiimide. Lorsque l'aminoacyle mis en œuvre possède une fonction acide supplémentaire (notamment dans le cas de l'acide glutamique), ces fonctions seront par exemple protégées, par des groupes t-butylester.
- 10 Dans le cas de la synthèse progressive, acide aminé par acide aminé, la synthèse débute de préférence par la condensation de l'amino-acide C-terminal avec l'aminoacide qui correspond à l'aminoacyle voisin dans la séquence désirée et ainsi de suite, de proche en proche, jusqu'à l'acide aminé N-terminal. Selon une autre technique préférée de l'invention, on a recours à celle décrite par R.D. MERRIFIELD dans l'article intitulé "Solid phase peptide synthesis" (J. Am. Soc., 45, 2149-2154).
- 15 Pour fabriquer une chaîne peptidique selon le procédé de MERRIFIELD, on a recours à une résine polymère très poreuse, sur laquelle on fixe le premier acide aminé C-terminal de la chaîne. Cet acide aminé est fixé sur la résine par l'intermédiaire de son groupe carboxylique et sa fonction amine est protégée, par exemple par le groupe t-butyloxycarbonyle.
- 20 Lorsque le premier acide aminé C-terminal est ainsi fixé sur la résine, on enlève le groupe protecteur de la fonction amine en lavant la résine avec un acide.
- Dans le cas où le groupe protecteur de la fonction amine est le groupe t-butyloxycarbonyle, il peut être éliminé par traitement de la résine à l'aide d'acide trifluoroacétique.
- 25 On couple ensuite le deuxième acide aminé qui fournit le second amino-acyle de la séquence recherchée, à partir du résidu amino-acyle C-terminal sur la fonction amine déprotégée du premier acide aminé C-terminal fixé sur la chaîne. De préférence, la fonction carboxyle de ce deuxième acide aminé est activée, par exemple par la dicyclohexyl-carbodiimide, et la fonction amine est protégée, par exemple par le t-butyloxycarbonyle.
- 30 On obtient ainsi la première partie de la chaîne peptidique recherchée, qui comporte deux acide aminés, et dont la fonction amine terminale est protégée. Comme précédemment, on déprotège la fonction amine et on peut ensuite procéder à la fixation du troisième aminoacyle, dans les conditions analogues à celles de l'addition du deuxième acide aminé C-terminal.
- On fixe ainsi, les uns après les autres, les acides aminés qui vont constituer la chaîne peptidique sur le groupe amine chaque fois déprotégé au préalable de la portion de la chaîne peptidique déjà formée, et qui est rattachée à la résine.
- 35 Lorsque la totalité de la chaîne peptidique désirée est formée, on élimine les groupes protecteurs des différents acide aminés constituant la chaîne peptidique et on détache le peptide de la résine par exemple à l'aide d'acide fluorhydrique.
- 40 L'invention concerne également les oligomères hydrosolubles des peptides monomères sus-indiqués. L'oligomérisation peut provoquer un accroissement de l'immunogénicité des peptides monomères selon l'invention. Sans qu'une telle indication chiffrée puisse être considérée comme limitative, on mentionnera néanmoins que ces oligomères peuvent, par exemple, contenir de 2 à 10 unités monomères.
- Les unités monomères entrant dans cet oligomère sont soit toutes constituées par le polypeptide de séquence 1 ou par le polypeptide de séquence 2, soit par l'un et l'autre de ces polypeptides.
- 45 On peut avoir recours, pour réaliser l'oligomérisation, à toute technique de polymérisation couramment utilisée dans le domaine des peptides, cette polymérisation étant conduite jusqu'à l'obtention d'un oligomère ou polymère contenant le nombre de motifs monomères requis pour l'acquisition de l'immunogénicité désirée.
- Une méthode d'oligomérisation ou de polymérisation du monomère consiste dans la réaction de celui-ci avec un agent de réticulation tel que le glutaraldéhyde.
- 50 On peut également avoir recours à d'autres méthodes d'oligomérisation ou de couplage, par exemple à celle mettant en jeu des couplages successifs d'unités monomères, par l'intermédiaire de leurs fonctions terminales carboxyle et amine en présence d'agents de couplage homo- ou hétéro- bifonctionnels.
- On peut également pour la production de molécules comportant un ou plusieurs motifs de 17 acides aminés tels que définis ci-dessus, avoir recours à des techniques du génie génétique mettant en œuvre des micro-organismes transformés par un acide nucléique déterminé comprenant des séquences nucléotidiques appropriées correspondantes.
- 55 L'invention concerne également les acides nucléiques contenant une ou plusieurs séquences issues de la séquence de l'ADNc du virus HIV-2 ROD. Ces séquences repérées par la numérotation figurant sur la séquence précédemment décrite, codent pour certains peptides intéressants de l'invention.

Séquence codant pour env1 nucléotides 7850 à 7927

	"	"	<u>env2</u>	"	8030 à 8095
5	"	"	<u>env3</u>	"	7601 à 7636
	"	"	<u>env4</u>	"	6170 à 6247
	"	"	<u>env5</u>	"	6294 à 6349
10	"	"	<u>env6</u>	"	6392 à 6445
	"	"	<u>env7</u>	"	6724 à 6763
	"	"	<u>env8</u>	"	6794 à 6838
	"	"	<u>env9</u>	"	7112 à 7162
15	"	"	<u>env10</u>	"	7253 à 7336
	"	"	<u>env11</u>	"	7358 à 7426
	"	"	<u>gag1</u>	"	1535 à 1597

20

L'invention concerne enfin les acides nucléiques correspondants du virus SIV, contenant une ou plusieurs séquences issues de l'ADNc du virus SIV-1. Ces séquences codant pour les peptides env1 à env11 et gag1 peuvent être repérées sur la figure 3 par comparaison avec les séquences correspondantes décrites pour HIV-2.

Il va de soi que l'invention concerne les acides nucléiques correspondant à des séquences placées en des régions analogues des ADNc dérivés de variants de HIV-2 ROD ou de SIV, ainsi que tous les acides nucléiques dont les modifications vis à vis des précédents résulteraient de la mise à profit de la dégénérescence du code génétique.

L'invention concerne encore les conjugués obtenus par couplage covalent des peptides selon l'invention (ou des susdits oligomères) à des molécules porteuses (naturelles ou synthétiques), physiologiquement acceptables et non toxiques, par l'intermédiaire de groupements réactifs complémentaires respectivement portés par la molécule porteuse et le peptide. Des exemples de groupements appropriés sont illustrés dans ce qui suit :

A titre d'exemple de molécules porteuses ou supports macromoléculaires entrant dans la constitution des conjugués selon l'invention, on mentionnera des protéines naturelles, telles que l'anatoxine tétanique, l'ovalbumine, des sérum albumines, des hémocyanines, etc...

A titre de support macromoléculaires synthétiques, on mentionnera par exemple des polylysines ou des poly(D-L-alanine)-poly(L-lysine).

La littérature mentionne d'autres types de supports macromoléculaires susceptibles d'être utilisés, lesquels présentent en général un poids moléculaire supérieur à 20 000.

Pour synthétiser les conjugués selon l'invention, on peut avoir recours à des procédés connus en soi, tels que celui décrit par FRANTZ et ROBERTSON dans Infect. and Immunity, 33, 193-198 (1981), ou celui décrit dans Applied and Environmental Microbiology, (octobre 1981), vol. 42, n° 4, 611-614 par P.E. KAUFFMAN en utilisant le peptide et la molécule porteuse appropriée.

Dans la pratique, on utilisera avantageusement comme agent de couplage les composés suivants, cités à titre non limitatif : aldéhyde glutarique, chloroformiate d'éthyle, carbodiimides hydrosolubles [N-éthyl-N'(3-diméthylaminopropyl) carbodiimide, HCl], diisocyanates, bis-diazobenzidine, di- et trichloro-s-triazines, bromures de cyanogène, ainsi que les agents de couplage mentionnés dans Scand. J. Immunol., (1978), vol. 8, p. 7-23 (AVRAMEAS, TERNYNCK, GUESDON).

On peut avoir recours à tout procédé de couplage faisant intervenir d'une part une ou plusieurs fonctions réactives du peptide et d'autre part, une ou plusieurs fonctions réactives de molécules supports. Avantageusement, il s'agit des fonctions carboxyle et amine, lesquelles peuvent donner lieu à une réaction de couplage en présence d'un agent de couplage du genre de ceux utilisés dans la synthèse des protéines, par exemple, le 1-éthyl-3-(3-diméthylaminopropyl) carbodiimide, le N-hydroxybenzotriazole, etc... On peut encore avoir recours à la glutaraldéhyde, notamment lorsqu'il s'agit de relier entre eux des groupes aminés respectivement portés par le peptide et la molécule support.

Les peptides selon l'invention possèdent des propriétés antigéniques. Ils peuvent donc être utilisés dans des procédés de diagnostic pour la détection d'une infection par le virus HIV-2.

Comme on l'a déjà mentionné, des études ont permis de distinguer deux groupes de peptides pouvant être mis en œuvre dans des procédés de détection d'anticorps contre le virus HIV-2 dans un fluide biologique humain, notamment un sérum ou un liquide céphalo-rachidien.

Un premier groupe (I) comprend les peptides gag1. Ces peptides reconnaissent des anticorps anti-HIV-2 et sont

**EP 0 750 041 A2**

dono capables de détecter une infection par HIV-2. Ils reconnaissent également dans une certaine mesure des anticorps anti-HIV-1.

Un second groupe (II) comprend des peptides qui correspondent plus particulièrement à ceux qui sont situés dans la partie transmembranaire et dans la fin de la partie externe de la protéine d'enveloppe. Ces peptides sont ceux précédemment désignés par env1, env2 et env3. Ils permettent la reconnaissance spécifique de la présence d'anticorps contre HIV-2 et permettent donc de discriminer chez une personne les infections passées ou présentes dues à un HIV, plus particulièrement entre celles qui ont été provoquées par un HIV-2 et celles qui l'ont été par un HIV-1.

L'invention concerne également une composition contenant au moins l'un des susdits peptides ou au moins un oligomère de ce peptide, caractérisée en ce qu'elle a la capacité d'être reconnue par des sérum s d'origine humaine 10 contenant des anticorps contre le virus HIV-2.

L'invention concerne un procédé de diagnostic in vitro un ou des peptides selon l'invention pour la détection d'anticorps contre HIV-2 dans des fluides biologiques, en particulier dans des sérum humains.

D'une façon générale le procédé de diagnostic in vitro ci-dessus comprend les étapes suivantes :

- 15 - la mise en contact de ce liquide biologique avec lesdits peptides,  
- la détection de la présence éventuelle d'un complexe peptide-anticorps par des méthodes physiques ou chimiques, dans ledit liquide biologique.

Dans un mode de réalisation préféré de l'invention, la détection du complexe antigène-anticorps est réalisée grâce 20 à des tests immunoenzymatiques (du type ELISA), immunofluorescents (du type IFA), radioimmunologiques (du type RIA) ou des tests de radioimmunoprécipitation (du type RIPA).

Ainsi l'invention concerne également tout peptide selon l'invention marqué à l'aide d'un marqueur adéquat du type enzymatique, fluorescent, radioactif, etc...

De telles méthodes comprennent par exemple les étapes suivantes :  
25

- dépôt de quantités déterminées d'une composition peptidique selon l'invention dans les puits d'une microplaques de titration,
- introduction dans lesdits puits de dilutions croissantes du sérum devant être diagnostiqués,
- incubation de la microplaques,
- rinçages répétés de la microplaques,
- introduction dans les puits de la microplaques d'anticorps marqués contre des immunoglobulines du sang, le marquage de ces anticorps ayant été réalisé à l'aide d'une enzyme sélectionnée parmi celles qui sont capables d'hydrolyser un substrat en modifiant l'absorption des radiations de ce dernier, au moins à une longueur d'onde déterminée,
- 35 - détection, en comparaison avec un témoin de contrôle, de la quantité de substrat hydrolysé.

L'invention concerne également des coffrets ou kits pour le diagnostic in vitro de la présence d'anticorps contre les virus HIV-2 et, dans certains cas, HIV-1 dans un milieu biologique qui comprennent ;

- 40 - une composition peptidique selon l'invention,  
- les réactifs pour la constitution du milieu propice à la réalisation de la réaction immunologique,  
- les réactifs permettant la détection du complexe antigènes-anticorps produit par la réaction immunologique. De tels réactifs peuvent également porter un marqueur, ou être susceptibles d'être reconnus à leur tour par un réactif marqué. Plus particulièrement dans le cas où la composition polypeptidique sus-mentionnée n'est pas marquée.  
45 - un tissu fluide biologique de référence dépourvu d'anticorps reconnus par la composition polypeptidique sus-mentionnée,

L'invention concerne les anticorps eux-mêmes formés contre les peptides de l'invention.

Il va de soi que cette production n'est pas limitée aux anticorps polyclonaux.  
50 Elle s'applique encore à tout anticorps monoclonal produit par tout hybridome susceptible d'être formé, par des méthodes classiques, à partir des cellules spléniques d'un animal, notamment de souris ou de rat, immunisés contre l'un des peptides de l'invention, d'une part et des cellules d'une lignée de cellule myélome approprié d'autre part, et d'être sélectionné, par sa capacité à produire des anticorps monoclonaux reconnaissant le peptide initialement mis en œuvre pour l'immunisation des animaux.

55 L'invention concerne également des compositions immunogènes pour la production de vaccins dont le principe actif est constitué par au moins un peptide selon l'invention, ou un oligomère de ce peptide, ou un peptide sous forme conjuguée avec une molécule porteuse, caractérisées en ce qu'elles induisent la production d'anticorps contre les susdits peptides en quantité suffisante pour aussi inhiber les protéines du rétrovirus HIV-2, voire même le rétrovirus

**EP 0 750 041 A2**

HIV-2 entrant en association avec un véhicule pharmaceutiquement acceptable.

Les compositions immunogènes pour la production de vaccins comprennent de façon avantageuse plus particulièrement au moins l'un des peptides précédemment désignés par env4, env5, env6, env7, env8, env9, env10, env11 voir des mélanges de ceux-ci.

5 Parmi ces peptides aptes à constituer des principes actifs de vaccins certains sont particulièrement préférés car ils possèdent une structure de base en acides aminés correspondant à des régions des glycoprotéines d'enveloppe qui présentent un important degré de conservation, non seulement dans les HIV-2, et dans les SIV, mais également dans les HIV-1. Ces peptides particulièrement préférés sont les peptides désignés par env4, certains peptides env5, env6 et env10.

10 Dans un mode de réalisation préféré de l'invention les peptides immunogènes (ou fragments de ces peptides) aptes à constituer des principes actifs de vaccins sont choisis parmi ceux dont les formules correspondent à des séquences qui, dans les glycoprotéines d'enveloppe de HIV-2, SIV et HIV-1 présentant une homologie en acides aminés supérieure à 50%, qui appartiennent à la partie externe de l'enveloppe du virus, qui sont dépourvus ou presque de délétions, et qui renferment des résidus de cystéine favorables à la stabilisation des liaisons et à la constitution de boucles d'ancre.

15 Les peptides suivants appartiennent à cette catégorie de peptides préférés.

**env4**

20 **XVTV-YGVP-W--ATZ**

**env5**

25 **XL-NVTE-FZ**

**env6**

30 **XKPCVKL-PLC-Z**

**env7**

35 **XN-S-I-Z**

**env10**

40 **XNC-GEF-YC-Z**

**env11**

**XC-I-Q-IZ**

45

Des compositions pharmaceutiques avantageuses sont constituées par des solutions, suspensions ou liposomes injectables contenant une dose efficace d'au moins un produit selon l'invention. De préférence, ces solutions, suspensions ou liposomes sont réalisés dans une phase aqueuse stérilisée isotonique, de préférence saline ou glucosée.

50 L'invention concerne plus particulièrement de telles suspensions, solutions ou forme liposome qui sont aptes à être administrées par injections intradermiques, intramusculaires ou sous-cutanées, ou encore par scarifications.

Elle concerne également des compositions pharmaceutiques administrables par d'autres voies, notamment par voie orale.

55 Les compositions pharmaceutiques selon l'invention, utilisables en tant que vaccins pour être efficaces dans la production d'anticorps contre le virus HIV-2, peuvent à titre d'exemple être administrées à des doses situées entre 10 et 500 µg/kg, de peptides selon l'invention, de préférence de 50 à 100 µg/kg.

Ces doses sont citées à titre d'exemple et ne possèdent en aucun cas un caractère limitatif.

Comme on l'a déjà indiqué plus haut les différents peptides qui ont été définis peuvent comprendre des modifications qui n'ont pas pour effet de modifier de façon fondamentale leurs propriétés immunologiques. Les peptides équi-

**EP 0 750 041 A2**

- valents qui en résultent entrent dans le champ des revendications qui suivent. A titre d'exemples de peptides équivalents on mentionnera ceux dont les structures en correspondance avec des régions des ADNc d'autres variants de HIV-2 de SIV ou de HIV-1, lorsque ces régions ont été mises en alignement dans des conditions semblables à celles qui ont été évoquées ci-dessus, à propos de HIV-2 ROD, SIV et HIV-1 BRU. A titre d'autres de ces peptides, on 5 mentionnera ceux dont les structures sont en correspondance avec de telles régions dans les ADNc qui ont fait l'objet de dépôts à la CNCM, notamment sous les numéros I-502, I-642 (HIV-2 IRMO), I-643 (HIV-2 EHO) ainsi que, dans les cas appropriés, des variants de HIV-1 qui ont fait l'objet de dépôts à la CNCM sous les numéros I-232, I-240, I-241, I-550, I-551.
- Les peptides selon l'invention peuvent encore être définis par les formules suivantes (dans lesquels X, Z et les 10 tirets "-" ont les significations sus-indiquées) :

15

20

25

30

35

40

45

50

55

XRV-AIEKYL-DQA-LN-WGCAFRQVCZ  
XAIEKYL-DZ

5  
X-LE-AQIQQEKNMYELQKLNSWZ  
XQIQQEKNZ

10  
XELGDYKLVEITPIG-APT--KR----Z  
XYKLVEITPIG-APT--KRZ

15  
X----VTV-YGVP-W--AT--LFCA-Z  
XVTV-YGVP-W--ATZ

20  
X----E--L-NVTE-F--W-NZ  
XL-NVTE-FZ

25  
XL---S-KPCVKL-PLC---Z  
XKPCVKL-PLC-Z  
XS-KPCVKL-PLC-Z

30  
X---N-S-I---C-Z  
XN-S-I-Z

35  
XYC-P-G-A-L-C-N-TZ

40  
X-----A-C-----W--Z  
NKRPRQAWCWFKG-KWKD

45  
X-G-DPE-----NC-GEF-YC-----NZ

X-----C-I-Q-I-----G---YZ

50  
L'invention concerne également outre les peptides de SIV déjà décrits, les protéines codées par l'ADNc du virus SIV. Elle concerne également les protéines de tout virus immunologiquement étroitement apparenté à SIV-1mac, en particulier tout virus dont les protéines et les glycoprotéines d'enveloppe croisent immunologiquement et dont les ADNc présentent un pourcentage d'homologie d'au moins 95% et de préférence d'au moins 98%.

55  
En particulier l'invention concerne :

- 1/ les protéines et glycoprotéines de l'enveloppe codées par le gène env et représentées à la figure 3,
- 2/ la protéine GAG représentée à la figure 4,

- 3/ la protéine POL représentée à la figure 5,  
 4/ la protéine Q représentée à la figure 6,  
 5/ la protéine R représentée à la figure 7,  
 6/ la protéine X représentée à la figure 8,  
 5  
 7/ la protéine F représentée à la figure 9,  
 8/ la protéine TAT représentée à la figure 10,

Les acides aminés des protéines précitées de SIV, ont été représentées en alignement avec les séquences d'acides aminés des protéines correspondantes du virus HIV-2 les points verticaux figurant entre les deux séquences 10 correspondent aux acides aminés communs entre les protéines des deux virus.

Les séquences d'ADNc codant pour les protéines précitées apparaissent sur la figure 1B. L'invention concerne, 15 outre les séquences nucléiques précitées toute séquence nucléique modifiée, qui code également pour les protéines du rétrovirus SIV ou d'un variant.

Ces séquences d'ADNc repérées par la numérotation figurant sur les séquences décrites précédemment (figure 1B) sont les suivantes :

- séquence codant pour <u>GAG</u> , nucléotides 551 à 2068				
20	-	"	"	<u>POL</u> ,
	-	"	"	<u>Q</u> ,
	-	"	"	<u>X</u> ,
	-	"	"	<u>R</u> ,
25	-	"	"	<u>F</u> ,
	-	"	"	<u>TAT-1</u>
	-	"	"	<u>ART-1</u>
30	-	"	"	<u>TAT-2</u>
	-	"	"	<u>ART-2</u>
	-	"	"	<u>ENV</u>
				" 1726 à 4893
				" 4826 à 5467
				" 5298 à 5633
				" 5637 à 5939
				" 8569 à 9354
				" 5788 à 6084
				" 6014 à 6130
				" 8296 à 8391
				" 8294 à 8548
				" 6090 à 8732

35 L'invention concerne donc naturellement les protéines précédemment décrites, lorsqu'elles sont obtenues à partir du virus SIV ou lorsqu'elles sont préparées par une méthode de synthèse, notamment par l'une des méthodes déjà citées en rapport avec la synthèse des peptides de plus petite taille.

L'invention concerne également l'utilisation des protéines précédentes pour le diagnostic de la présence éventuelle 40 d'anticorps dirigés contre les protéines de HIV-2, voire contre HIV-2 en entier, ou pour certaines d'entre elles l'utilisation aux fins de diagnostic d'une infection due à l'un des virus HIV. Ainsi le peptide GAG codé par le gène correspondant peut être utilisé pour repérer la présence éventuelle d'anticorps anti-HIV-1 ou anti-HIV-2. Les protéines ENV sont utilisées de préférence pour le diagnostic spécifique d'une infection due à HIV-2 ou un de ses variants, parfois pour le diagnostic d'une infection par HIV-2 ou HIV-1.

45 L'invention concerne donc également un procédé de diagnostic in vitro de détection d'anticorps contre HIV-2 et éventuellement contre HIV-1 dans des fluides biologiques et en particulier dans des sérums humains. De tels procédés applicables pour l'utilisation des protéines précédentes de SIV comme protéines de diagnostic, ont déjà été décrits dans la présente invention.

L'invention concerne aussi des coffrets ou "kits" pour le diagnostic in vitro de la présence d'anticorps le virus HIV-50 2 et dans certains cas contre HIV-1 dans un milieu biologique. De tels kits mettant en oeuvre les peptides précédents ont également été décrits dans la présente invention.

L'invention concerne également des compositions immunogènes pour la production de vaccins, dont le principe actif est constitué de façon avantageuse par au moins la partie de la protéine ENV du virus SIV, cette protéine pouvant être sous forme conjuguée avec une molécule porteuse. Ces compositions immunogènes induisent la production d'anticorps contre le susdit peptide en quantité suffisante pour inhiber les protéines du rétrovirus HIV-2, voire le rétrovirus HIV-2 lui-même.

Toutefois l'utilisation aux fins de diagnostic des protéines de SIV n'est en rien limitée à celle des seuls protéines ENV ou GAG. D'autres protéines parmi celles décrites peuvent être envisagées, pour préparer des compositions de

diagnostic voire de vaccin.

**Revendications**

5. 1. Séquence de nucléotides caractérisée en ce qu'elle répond à la séquence nucléotidique représentée à la figure 1B ou à la figure 1C, ou en ce qu'elle contient la séquence nucléotidique représentée à la figure 1B ou à la figure 1C, ou en ce qu'il s'agit d'une partie de la séquence représentée à la figure 1B ou à la figure 1C, ladite partie de séquence codant pour un peptide reconnu par des anticorps présents dans le sérum d'un patient infecté par un rétrovirus HIV-2 ou étant utilisable comme sonde pour la détection dans un échantillon biologique, de la présence d'un rétrovirus HIV-2.

10. 2. Séquence de nucléotides selon la revendication 1, caractérisée en ce qu'elle comprend l'une des séquences suivantes identifiées dans la figure 1B ou dans la figure 1C.

15	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="padding: 2px;"><u>GAG</u> s'étendant entre les nucléotides</td><td style="padding: 2px; text-align: right;">550 à 2068</td></tr> <tr> <td style="padding: 2px;"><u>POL</u></td><td style="padding: 2px; text-align: right;">1726 à 4893</td></tr> <tr> <td style="padding: 2px;">Q</td><td style="padding: 2px; text-align: right;">4826 à 5467</td></tr> <tr> <td style="padding: 2px;">X</td><td style="padding: 2px; text-align: right;">5298 à 5633</td></tr> <tr> <td style="padding: 2px;">R</td><td style="padding: 2px; text-align: right;">5637 à 5939</td></tr> <tr> <td style="padding: 2px;">F</td><td style="padding: 2px; text-align: right;">8569 à 9354</td></tr> <tr> <td style="padding: 2px;">TAT-1</td><td style="padding: 2px; text-align: right;">5788 à 6084</td></tr> <tr> <td style="padding: 2px;">ART-1</td><td style="padding: 2px; text-align: right;">6014 à 6130</td></tr> <tr> <td style="padding: 2px;">TAT-2</td><td style="padding: 2px; text-align: right;">8296 à 8391</td></tr> <tr> <td style="padding: 2px;">ART-2</td><td style="padding: 2px; text-align: right;">8294 à 8548</td></tr> <tr> <td style="padding: 2px;">LTR</td><td style="padding: 2px; text-align: right;">8950 à 9468 et 1 à 316</td></tr> <tr> <td style="padding: 2px;">ENV</td><td style="padding: 2px; text-align: right;">6090 à 8732</td></tr> </table>	<u>GAG</u> s'étendant entre les nucléotides	550 à 2068	<u>POL</u>	1726 à 4893	Q	4826 à 5467	X	5298 à 5633	R	5637 à 5939	F	8569 à 9354	TAT-1	5788 à 6084	ART-1	6014 à 6130	TAT-2	8296 à 8391	ART-2	8294 à 8548	LTR	8950 à 9468 et 1 à 316	ENV	6090 à 8732
<u>GAG</u> s'étendant entre les nucléotides	550 à 2068																								
<u>POL</u>	1726 à 4893																								
Q	4826 à 5467																								
X	5298 à 5633																								
R	5637 à 5939																								
F	8569 à 9354																								
TAT-1	5788 à 6084																								
ART-1	6014 à 6130																								
TAT-2	8296 à 8391																								
ART-2	8294 à 8548																								
LTR	8950 à 9468 et 1 à 316																								
ENV	6090 à 8732																								

20. 3. Séquence de nucléotides selon la revendication 1 ou la revendication 2, caractérisée en ce qu'elle comprend l'une des séquences suivantes identifiées dans la séquence de la figure 1A:
25. 35. - séquence correspondant aux nucléotides 7850 à 7927  
       - séquence correspondant aux nucléotides 8030 à 8095  
       - séquence correspondant aux nucléotides 7601 à 7636  
       - séquence correspondant aux nucléotides 6170 à 6247  
       - séquence correspondant aux nucléotides 6294 à 6349  
       - séquence correspondant aux nucléotides 6392 à 6445  
       - séquence correspondant aux nucléotides 6724 à 6763  
       - séquence correspondant aux nucléotides 6794 à 6838  
       - séquence correspondant aux nucléotides 7112 à 7162  
       - séquence correspondant aux nucléotides 7253 à 7336  
       - séquence correspondant aux nucléotides 7358 à 7426  
       - séquence correspondant aux nucléotides 1535 à 1597
40. 4. Séquence de nucléotides caractérisée en ce qu'il s'agit d'une séquence selon la revendication 1 modifiée par dégénérescence du code génétique.
45. 5. Séquence de nucléotides selon la revendication 1, caractérisée en ce qu'il s'agit de la séquence contenue dans le plasmide pSIV-1.1 (CNCM I-658) ou dans le plasmide pSIV-1.2 (CNCM 1-659).
50. 6. Séquence de nucléotides selon la revendication 1, caractérisée en ce qu'elle code pour un polypeptide choisi parmi
55.     POL<sub>ROD</sub> ou POL<sub>MAC</sub> représenté à la figure 5  
       Q<sub>ROD</sub> ou Q<sub>MAC</sub> représentée à la figure 6  
       R<sub>ROD</sub> ou R<sub>MAC</sub> représentée à la figure 7

**EP 0 750 041 A2**

$X_{ROD}$  ou  $X_{MAC}$  représentée à la figure 8  
 $F_{ROD}$  ou  $F_{MAC}$  représentée à la figure 9  
 $TAT_{ROD}$  ou  $TAT_{MAC}$  représentée à la figure 10  
 $ART_{ROD}$  ou  $ART_{MAC}$  représentée à la figure 11.

- 5      7. Séquence de nucléotides selon l'une quelconque des revendications 1 à 6, caractérisée en ce qu'elle est marquée.
8. Utilisation d'une séquence de nucléotides selon l'une quelconque des revendications 1 à 7 comme sonde pour la détection dans un échantillon biologique, d'une infection par un rétrovirus HIV-2.
- 10     9. Acide nucléique recombinant caractérisé en ce qu'il comprend une séquence de nucléotides selon l'une quelconque des revendications 1 à 6, insérée dans un acide nucléique provenant d'un vecteur.
- 15     10. Procédé de diagnostic de la présence ou non du virus HIV-2 ou d'un variant, dans des échantillons de sérums ou d'autres liquides ou tissus obtenus à partir de patients suspectés d'être porteurs du virus HIV-2 comprenant :
- au moins une étape d'hybridation conduite dans des conditions stringentes, par mise en contact de l'ADN de cellules de l'échantillon du patient suspect avec une sonde selon la revendication 7 sur une membrane appropriée,
  - le lavage de ladite membrane avec une solution assurant la conservation de ces conditions stringentes de l'hybridation,
  - la détection de la présence ou non du virus HIV-2 par une méthode d'immunodétection.
- 20     11. Procédé selon la revendication 10, caractérisé en ce que l'étape d'hybridation est conduite dans des conditions non stringentes et en ce que le lavage de ladite membrane est effectué avec une solution assurant la conservation des conditions non stringentes.
- 25

30

35

40

45

50

55

FIG. 1.A

HIV2.ROD

R →  
 GTCCGCTCTGGAGAGGGCTGGCAGATTGAGCCCTGGGAGGTTCTCTCCACCACTAGCAGG  
 TAGAGCCCTGGGTGTTCCCTGCTAGACTCTCACCAAGCACTTGCGGGTGTGGCAGACGG  
 100 R US  
 CCCACGCTTGCTTGCTTAAACCTCTTAATAAAAGCTGCCACTTAGAAGCAAGTTAAGT  
 CTGTGCTCCCATCTCTCTAGTCGCCGCCTGGTCATTGGTGTTCACCTGAGTAACAAGA  
 200  
 CCTGGTCTGTTAGGACCCTTGGCTTGGAAACCGAGCCAGGAAAATCCCTAGCAGG  
 300  
 TTGGCCCTGAACAGGGACTTGAGAAGACTGAGAAGCTTGAAACACCTGAGTGAAAG  
 GCACTAAGGGCGGCAGGAACAAACCACGACGGAGTGCTCCTAGAAAGCCGGGGCGAGG  
 400  
 TACCAAAGGCAGCGTGGAGCCGGAGGAGAACGCGCTGGCTAACGTAAGTACCTA  
 CACCAAAAAACTGTAGCCGAAAGGGCTTGCTATCTACCTTACATAGCTAGAAAGATTGTC  
 500  
 MetGlyAlaArgAsnSerValLeuArgGlyLysAlaAspGluLeuGluArgIle  
 CGAGATGGGCCGACAAACTCCGTCTTGAGAGGA4AAAAGCAGATGAATTAGAAAGAAT  
 600  
 ArgLeuArgProGlyGlyLysLysTyrArgLeuLysHisIleValTrpAlaAlaAsn  
 CAGGTTACGGCCCGGGAAAGAAAAACTACAGGCTAAAAACATATTGTGTGCCACCGAA  
 .  
 LysLeuAspArgPheGlyLeuAlaGluSerLeuLeuGluSerLysGluGlyCysGlnLys  
 TAAATTGGACAGATTGGATTAGCAGAGGCCCTGGACTCAAAAGAGGGCTTGTCAAAA  
 700  
 IleLeuThrValLeuAspProMetValProThrGlySerGluAsnLeuIysSerLeuPhe  
 AATTCTTACACTTTAGATCCAATGGTACCCACAGGTCAGAAAATTAAAAAGTCCTT  
 .  
 AsnThrValCysValIleTrpCysIleHisAlaGluGluLysValLysAspThrGluGly  
 TAATACTGTCTGGTCATTGGTCATACACGGCAGAAGAGAAAAGTGAACATACGAAAG  
 800  
 AlaLysGlnIleValArgArgBisLeuValAlaGluThrGlyThrAlaGluLysMetPro  
 AGCAAAACAAATACTGGGAGACATCTACTGGCACAAACAGGAACTCCACAGAAAATGCC

FIG. 1A

SerThrSerArgProThrAlaProSerSerGluLysGlyGlyAsnTyrProValGluHis  
 AACGACAACTAGACCAACAGCACCCTAGCGAGAAGGGAGGAAATTACCCAGTGGCAACA  
  
 ValGlyGlyAsnTyrThrHisIleProLeuSerProArgThrLeuAsnAlaTrpValLys  
 TGTAGCCCCAACTACACCCATATACCGCTGAGTCCCCGACCTAAATGCCTGGGTAAA  
    1000  
  
 LeuValGluGluLysLysPheGlyAlaGluValValProGlyPheGluAlaLeuSerGlu  
 ATTACTAGAGGAAAAAAAGTTCCGGGGCAGAACTAGTGCCAGGATTTCAGGCCTCTCAGA  
  
 GlyCysThrProTyrAspIleAsnGlnMetLeuAsnCysValGlyAspHisGlnAlaAla  
 AGGCTGCCACCCCTATGATATCAACCAAATGCTTAATTGTCTGGGCACCATCAAGCAGC  
    1100  
  
 MetGlnIleIleArgGluIleIleAsnGluGluAlaAlaGluTrpAspValGlnHisPro  
 CATCCAGATAATCAGGGAGATTATCAATGAGGAACCAGCAGAACATGGATGTGCAACATCC  
    1200  
  
 IleProGlyProLeuProAlaGlyGlnLeuArgGluProArgGlySerAspIleAlaGly  
 AATACCAGGCCCTTACCGCCGGGCAGCTAGAGAGCCAAGGGATCTGACATACCCAGG  
  
 ThrThrSerThrValGluGluGlnIleGlnTrpMetPheArgProGlnAsnProValPro  
 GACAACAAAGCACAGTAGAAGAACAGATCCAGTGGATTTAGGCCACAAAATCCTGTACC  
    1300  
  
 ValGlyAsnIleTyrArgArgTrpIleGlnIleGlyLeuGlnLysCysValArgMetTyr  
 AGTAGGAAACATCTATAGAAGATGGATCCAGATAGGATTCAGAAGTGTCTCAGGATGTA  
  
 AsnProThrAsnIleLeuAspIleLysGlnGlyProLysGluProPheGlnSerTyrVal  
 CAACCCGACCAACATCCTAGACATAAAACAGGGACCAAAAGGAGCCGTTCCAAAGCTATGT  
    1400  
  
 AspArgPheTyrLysSerLeuArgAlaGlnGlnThrAspProAlaValLysAsnTrpMet  
 AGATAGATTCTACAAAAGCTTGGGCAGAACACAGATCCAGCAGTGAAGAATIGGAT  
    1500  
  
 ThrGlnThrLeuLeuValGlnAsnAlaAsnProAspCysLysLeuValLeuLysGlyLeu  
 GACCCAAACACTGCTAGTACAAAATGCCAACCCAGACTCTAAATTAGTGTCTAAAAGGACT  
  
 GlyMetAsnProThrLeuGluGluMetLeuThrAlaCysGlnGlyValGlyGlyProGly  
 AGGGATGAACCTACCTTACAAAGAGATGCTGACCCCTGTCAGGGGGTAGGTGGCCAGG  
    1600  
  
 GlnLysAlaArgLeuMetAlaGluAlaLeuLysGluValIleGlyProAlaProIlePro  
 CCAGAAAGCTAGATTAAATGCCAGAGGCCCTGAAACAGGTCTAGGACCTGCCCTATCCC  
  
 PheAlaAlaAlaGlnGlnArgLysAlaPheLysCysTrpAsnCysGlyLysGluGlyHis  
 ATTCCGACAGCCCCAGCAGAGAAAGGCTTAAATGCTGCACTGTGCAAAAGGAAGGGCA  
    1700  
  
 SerAlaArgGlnCysArgAlaProArgArgGlnGlyCysTrpLysCysGlyLysProGly  
 CTCGGCAAGACAATGCCGAGCACCCTAGAACCCAGGGCTGCTGGAAGTGTGGTAAGCCAGG  
    1800  
  
 ThrGlyArgPhePheArgThrGlyProLeuGly  
 HisIleMetThrAsnCysProAspArgGlnAlaGlyPheLeuGlyLeuGlyProTrpGly  
 ACACATCATGACAAACTGCCAGATAGACAGGCAGGTTTTAGGACTGGGCCCTGGGG  
  
 LysGluAlaProGlnLeuProArgGlyProSerSerAlaClyAlaAspThrAsnSerThr  
 LysLysProArgAsnPheProValAlaGlnValProGlnGlyLeuThrProThrAlaPro  
 AAACAAGCCCCGCAACTTCCCGTGGCCCAAGTTCCGCAGGGCTGACACCAACAGCACC  
    1900  
  
 ProSerGlySerSerSerGlySerThrGlyGluIleTyrAlaAlaArgGluLysThrGlu  
 ProValAspProAlaValAspLeuLeuGluLysTyrMetGlnGlnGlyLysArgGlnArg  
 CCCACTGGATCCAGCAGTGGATCTACTGGAGAAATATATGCAAGCAAGGGAAAAGACAGAC  
  
 ArgAlaGluArgGluThrIleGlnGlySerAspArgGlyLeuThrAlaProArgAlaGly  
 GluGlnArgGluArgProTyrLysGluValThrGluAspLeuLeuHisLeuGluGlnGly  
 AGAGCAGACAGAGAGACCATAAAGGAAGTGAACAGAGGACTTACTGCACCTCGAGCAGGG  
 (fig. 1A-suite k)

GlyAspThrIleGlnGlyAlaThrAsnArgGlyLeuAlaAlaProGlnPheSerLeuTrp  
 GluThrProTyrArgGluProProThrGluAspLeuLeuHisLeuAsnSerLeuPheGly  
 GGAGACACCATAACAGGGAGCCACCAACACAGGACTTGCTGCACCTCAATTCTCTTTGG  
 2100  
 LysArgProValValThrAlaTyrIleGluGlyGlnProValGluValLeuLeuAspThr  
 LysAspGln  
 AAAAGACCACTAGTCACACCATACATTGACGGTCAGCCAGTAGAAGTCTTGTAGACACA  
 GlyAlaAspAspSerIleValAlaGlyIleGluLeuGlyAsnAsnTyrSerProLysIle  
 GGGGCTGACCACTAAATAGTAGCAGGAATAGAGTTAGGGAAATAATAGCCCCAAAAATA  
 2200  
 ValGlyGlyIleGlyGlyPheIleAsnThrLysGluTyrLysAsnValGluIleGluVal  
 GTAGGGGAATAGGGGATTCTAAATACCAAGGAATAAAAATGAGAAATAGAAAGTT  
 LeuAsnLysValArgAlaThrIleMetThrGlyAspThrProIleAsnIlePheGly  
 CTAATAAAAGGTACGGGCCACCATATGACAGGGCACCCCCATCAACATTTTGGC  
 2300  
 ArgAsnIleLeuThrAlaLeuGlyMetSerLeuAsnLeuProValAlaLysValGluPro  
 AGAAATATTCTGACAGCCTTACGGCATGTCATTAAATCTACCGTCGCCAAAGTAGAGCCA  
 2400  
 IleLysIleMetLeuLysProGlyLysAspGlyProLysLeuArgGlnTrpProLeuThr  
 ATAAAAATAATGCTAAAGCCACGGAAACATGGACCAAAACTGAGACAATGCCCTTAACA  
 LysGluLysIleGluAlaLeuLysGluIleCysGluLysMetGluLysGluGlyGluLeu  
 AAAGAAAAAAATAGAACGACTAAAAGAAATCTGTGAAAAAAATGCAAAAGAACGGCAGCTA  
 2500  
 2500  
 GluGluAlaProProThrAsnProTyrAsnThrProThrPheAlaIleLysLysAsp  
 GAGGAAGCACCTCAACTAACCTTATAATACCCCCACATITGCAATCAAGAAAAAGGAC  
 LysAsnLysTrpArgMetLeuIleAspPheArgGluLeuAsnLysValThrGluAspPhe  
 AAAAACAAATGGAGGATGCTAATAGATTTCAAGAGAACTAAACAAGGTAACTCAAGATTTC  
 2600  
 ThrGluIleGlnLeuGlyIleProHisProAlaGlyLeuAlaLysLysArgArgIleThr  
 ACAGAAATTCTAGTTAGGAATCCCACACCCACCGGGTGGCCAAGAACAGAACAAATTACT  
 2700  
 ValLeuAspValGlyAspAlaTyrPheSerIleProLeuHisGluAspPheArgProTyr  
 CTACTACATCTAGGGATGCTTACTTTCCATACCACTACATGAGGACTTTAGACCATAT  
 ThrAlaPheThrLeuProSerValAsnAsnAlaGluProGlyLysArgTyrIleTyrLys  
 ACTGCATTTACTCTACCATCAGTGAACAATGCAGAACCGGAAAGATAACATATATAAA  
 2800  
 ValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnHisThrMetArgGlnVal  
 GTCTTGCCACAGGGATGGAAGGGATCACCAAGCAATTTCAACACACAATGAGACAGGTA  
 LeuGluProPheArgLysAlaAsnLysAspValIleIleIleGluTyrMetAspAspIle  
 TTAGAACCATTCAGAAAAACAAACAGGATGTCATTATCATTCACTGAGGATGATGATATC  
 2900  
 LeuIleAlaSerAspArgThrAspLeuGluHisAspArgValValLeuGlnLeuLysGlu  
 TTAATAGCTACTGACACGGACAGATTTACAAACATGATAGGGTACTCCTGCAGCTCAACCAA  
 3000  
 LeuLeuAsnGlyLeuGlyPheSerThrProAspGluLysPheGlnLysAspProProTyr  
 CTTCTAAATGCCCTACGATTTCTACCCACATGAGAACTTCAAAAGACCCCTCCATAC  
 HisTrpMetGlyTyrGluLeuTrpProThrLysTrpLysLeuGlnLysIleGlnLeuPro  
 CACTGGATGGCTATGAACTATGGCAACTAAATGGAAGITGCAGAAAAATACAGTTGCC  
 3100  
 GlnLysGluIleTrpThrValAsnAspIleGlnLysLeuValGlyValLeuAsnTrpAla  
 CAAAAAGAAATATGGCAGACTCAATGACATCCAGAAGCTACTGGCTGTCCTAAATTGGGCA

(fig.1A-suite 2)

AlaGlnLeuTyrProGlyIleLysThrLysHisLeuCysArgLeuIleArgGlyLysMet  
 GCACAACCTCACCCAGGGATAAAAGACCAAAACACTTATCTAGGTTAACAGAGGAAAAATG  
 3200  
 ThrLeuThrGluGluValGlnTrpThrGluLeuAlaGluAlaGluLeuGluAsnArg  
 ACACCTCACAGAAGACTACAGTGGACAGAATTACCACACCAGCCTAGAACAAAAACAGA  
 3300  
 IleIleLeuSerGlnGluGlnGluGlyHisTyrTyrGlnGluGluLysGluLeuGluAla  
 ATTATCCTAACGCCAGGAACAAAGAGGGACACTATTACCAAGAACAGAAAAAGAGCTAGAACCA  
  
 TyrValClnLysAspGluGluAsnGluTrpThrTyrLysIleHisGlnGluGluLysIle  
 AGAGTCACAAAGGATCAAGAGAAATGAGTGGACATATAAAATACACCAGGAACAAAAAATT  
  
 LeuLysValGlyLysTyrAlaLysValLysAsnThrHisThrAsnGlyIleArgLeuLeu  
 CTAAAAGTAGGAAATATGCAAAGGTGAAAAACACCCATACCAATCGAACATTGTTA  
  
 AlaGlnValValGlnLysIleGlyLysGluAlaLeuValIleTrpGlyArgIleProLys  
 GCACAGGTAGTTCAAGAAATAGGAAAGAACGACTAGTCATTGGGACGAATACCAAAA  
 3500  
 PheHisLeuProValGluArgGluIleTrpGluGlnTrpTrpAspAsnTyrTrpGlnVal  
 TTTCACCTACCAGTAGAGAGAACAAATCTGGGAGCAGTGGTGGGATAACTACTGCCAACTG  
 3600  
 ThrTrpIleProAspTrpAspPheValSerThrProProLeuValArgLeuAlaPheAsn  
 ACATGGATCCCAGACTGGGACTTCGTGTCTACCCCACACTGGTCAGGTTAGCGTTAAC  
  
 LeuValGlyAspProIleProGlyAlaGluThrPheTyrThrAspGlySerCysAsnArg  
 CTGGTAGGGATCTTACACAGGTGCAGAGACCTCTACACAGATGGCATCTGCAATAGG  
 3700  
 GlnSerLysGluGlyLysAlaGlyTyrValThrAspArgGlyLysAspLysValLysLys  
 CAATCAAAACAAAGAAAACCACGGATATGTAACAGATAGAGGGAAAGACAAGGIAAAAGAAA  
  
 LeuGluGlnThrThrAsnGlnGlnAlaGluLeuGluAlaPheAlaMetAlaLeuThrAsp  
 CTAGACCAAACATACCAATCAGCAAGCAGAACTAGAACGCCTTGCGATGGCACTAACAGAC  
 3800  
 SerGlyProLysValAsnIleIleValAspSerGlnTyrValMetGlyIleSerAlaSer  
 TCGGGTCCAAAAGTTAATATTATAGACTCACAGTATGTAATGGGATCAGTCCAAGC  
 3900  
 GlnProThrGluSerGluSerLysIleValAsnGlnIleIleGluGluMetIleLysLys  
 CAACCAAACAGAGTCAGAAACTAAATAGTCAACAGATCATAGAAAGAAATGATAAAAAG  
  
 GluAlaIleTyrValAlaTrpValProAlaHisLysGlyIleGlyGlyAsnGlnGluVal  
 GAAGCAATCTATGTCATGGGCCCCAGCCCACAAAGGCATAGGGGAAACCACGAACTA  
 4000  
 AspHisLeuValSerGlnGlyIleArgGlnValLeuPheLeuGluLysIleGluProAla  
 GATCATTTAGTCAGTCAGGGTATCAGACAACGTTGCTGGAAAAAAATACACCCCCCT  
  
 GlnGluGluHisGluLysTyrHisSerAsnValLysGluLeuSerHisLysPheGlyIle  
 CAGGAAGAACATGAAAATATCATAGCAATGTAAGAACACTGTCTCATAAAATTGGATA  
 4100  
 ProAsnLeuValAlaArgGlnIleValAsnSerCysAlaGluCysGlnLysGlyGlu  
 CCCAATTAGTGGCAAGGAAATAGTAAACTCATGTCATGGCCCAATGTCAACAGAAACGGGAA  
 4200  
 AlaIleHisGlyGlnValAsnAlaGluLeuGlyThrTrpGlnMetAspCysThrHisLeu  
 GCTATACATGGGCAAGTAAATGCAAGACTAGGCACCTGGCAAATGGACTGCACACATTAA  
  
 GluGlyLysIleIleIleValAlaValHisValAlaSerGlyPheIleGluAlaGluVal  
 GAAGGAAAGATCATATAGTAGCAGTACATGTCAGTGGATTATAGAACGAGAACGTC  
 4300  
 IleProGlnGluSerGlyArgGlnThrAlaLeuPheLeuLeuLysLeuAlaSerArgTrp  
 ATCCCACAGGAATCAGGAAGAACAAACAGCAGTCATTGAAACTGGCAAGTAGGTGG  
 (fig.1A-suite 3)

ProIleThrHisLeuHisThrAspAsnGlyAlaAsnPheThrSerGlnGluValLysMet  
 CCAATAACACACTTGCATACACATAATGGTGCAACTTCACAGGAGGTAAAGATG  
                         4400  
 ValAlaTrpIleGlyIleGluGlnSerPheGlyValProTyrAsnProGlnSerGln  
 GTAGCATGGTGGATAGGTATACAACAATCCTTGAGTACCTTACAATCCACAGAGCCAA  
                         4500  
 GlyValValGluAlaMetAsnHisHisLeuLysAsnGlnIleSerGluThrIleValLeu  
 GGAGTAGTAGAAGCAATGAATCACCATCTAAAAAACCAAATAAGTGAAAACAATAGTACTA  
                         4600  
 MetAlaIleHisCysMetAsnPheLysArgArgGlyGlyIleGlyAspMetThrProSer  
 ATGGCAATTCAATTGCATGAATTAAAAAGAAGGGGGGAATAGGGATATGACTCCATCA  
                         4700  
 GluArgLeuIleAsnMetIleThrThrGluGlnGluIleGlnPheLeuGlnAlaLysAsn  
 GAAAGATTAATCAATATGATCACCAAGAACAGATAACAATTCTCCAAGGCCAAAT  
                         4800  
 SerLysLeuLysAspPheArgValTyrPheArgGluGlyArgAspGlnLeuTrpLysGly  
 TCAAAATTAAAAGATTTCCGGTCTATTCAGAGAAGGGCAGAGATCAGTTGTGAAACCGA  
                         4900  
 ProGlyGluLeuLeuTrpLysGlyGluGlyAlaValLeuValLysValGlyThrAspIle  
 CCTGGGAACTACTGTGGAAAGGAGAAGGAGCAGTCCTAGTCAACGGTACGGAAACAGACATA  
                         5000  
 LysIleIleProArgArgLysAlaLysIleIleArgAspTyrGlyGlyArgGlnGluMet  
                         MetGluGluAspLysArgTrp  
 AAAATAATACCAAGAAGGAAAGCCAAGATCATCAGAGACTATGGAGGAAGACAAGAGATC  
                         5100  
 AspSerGlySerHisLeuGluGlyAlaArgGluAspGlyGluMetAla  
                         IleValValProThrTrpArgValProGlyArgMetGluLysTrpHisSerLeuValLys  
 GATAGTGGTCCCACCTGGAGGGTGCAGGGAGGAAATGGCATAGCCTGTCAA  
                         5200  
 TyrLeuLysTyrLysThrLysAspLeuGluLysValCysTyrValProHisHisLysVal  
 GTATCTAAAATACAAAACAAAGGATCTAGAAAAGGTGTGCTATGTTCCCCACCATAAGGT  
                         5300  
 GlyTrpAlaTrpTrpThrCysSerArgValIlePheProLeuLysGlyAsnSerHisLeu  
 CCCATGGGCATGGTGGACTTGCACCCAGGGTAAATATTCCATTAAAAGGAAACAGTCATCT  
                         5400  
 GluIleGlnAlaTyrTrpAsnLeuThrProGluLysGlyTrpLeuSerSerTyrSerVal  
 AGAGATACAGGCATATTGCAACTAACACCAGAAAAGGATGGCTCTCCCTCTTATTCACT  
                         5500  
 ArgIleThrTrpTyrThrGluLysPheTrpThrAspValThrProAspCysAlaAspVal  
 AAGAATAACTTGGTACACAGAAAAGTTCTGGACAGATGTTACCCCAGACTGTGCAAGATGT  
                         5600  
 LeuIleHisSerThrTyrPheProCysPheThrAlaGlyGluValArgArgAlaIleArg  
 CCTAATACATAGCACTTATTCCCTTGCTTACAGCAGGTAAAGTAAGAACAGCCATCAG  
                         5700  
 GlyGluLysLeuLeuSerCysCysAsnTyrProArgAlaHisArgAlaGlnValProSer  
 AGGGGAAAGTATTGTCCTGCTGCAATTATCCCCGACCTCATAGACCCAGGTACCGTC  
                         5800  
 LeuGlnPheLeuAlaLeuValValValGlnGlnAsnAspArgProGlnArgAspSerThr  
                         MetThrAspProArgGluThrValPro  
 ACTTCAATTCTGGCCTTACTGGTACTGCAACAAAATGACAGACCCAGAGAGACAGTAC  
                         5900  
 ThrArgLysGlnArgArgAspTyrArgArgGlyLeuArgLeuAlaLysGlnAspSer  
                         ProGlyAsnSerGlyGluGluThrIleGlyGluAlaPheAlaTrpLeuAsnArgThrVal  
 CACCAAGGAAACAGCGCGAACAGACTATGGAGAGGGCTTCGGCTGGCTAAACAGGCACAG  
                         6000  
 ArgSerHisLysGlnArgArgSerSerGluSerProThrProArgThrTyrPheProGlyVal  
                         GluAlaIleAsnArgGluAlaValAsnHisLeuProArgGluLeuIlePheGlnValTrp  
 TAGAAGCCATAAACACAGAACGACTGAATCACCTACCCCCAGAACATTATTTCCAGGTGT  
 (fig.1A-suite 4)

AlaGluValLeuGluIleLeuAla  
 GlnArgSerTrpArgTyrTrpHisAspGluGlnGlyMetSerGluSerTyrThrLysTyr  
 CGCAGAGGTCCCTGGACATACTGGCATGATGAACAAAGGGATGTCAGAAACTTACACAAAGT  
 5500  
 ArgTyrLeuCysIleIleGlnLysAlaValTyrMetHisValArgLysGlyCysThrCys  
 ATAGATATTTCGCTATAATACAGAAAGCACTGTACATGCATGTTAGCAAAGGGTGACTT  
 LeuGlyArgGlyHisGlyProGlyGlyTrpArgProGlyProProProProProPro  
 GCCTGGGGAGGGGACATGGCCAGGAGGGTGGAGACCAGGGCTCCTCCTCCCCCTC  
 5600  
 MetAlaGluAlaProThrGluLeuProProValAspGlyThrProLeu  
 GlyLeuVal\*\*\*  
 CACGTCTGGCTAAATGGCTGAACCAACAGAGCTCCCCCGTGGATGGACCCACT  
 ArgGluProGlyAspGluTrpIleIleGluIleLeuArgGluIleLysGluGluAlaLeu  
 GAGGGAGCCAGGGATGAGTGGATAATAGAAATCTTGAGAGAAATAAAGAACAGCTT  
 LysHisPheAspProArgLeuIleAlaLeuGlyLysTyrIleTyrThrArgHisGly  
 MetGlu  
 AAAGCATTTCACCCCTCGCTTGCTAATTGCTTGGCAAATATATCTATACTAGACATGG  
 5800  
 AspThrLeuGluGlyAlaArgGluLeuIleLysValLeuGlnArgAlaLeuPheThrHis  
 ThrProLeuLysAlaProGluSerSerLeuLysSerCysAsnGluProPheSerArgThr  
 AGACACCCCTGAAGGCCAGAGAGCTATTAAACTCTGCAACGAGCCCTTTCACCGA  
 PheArgAlaGlyCysGlyHisSerArgIleGlyGlnThrArgGlyGlyAsnProLeuSer  
 SerGluGlnAspValAlaThrGlnGluLeuAlaArgGlnGlyGluGluIleLeuSerGln  
 CTTCAGACCCACGATCTGGCCACTCAAGAATTGGCAGACAAGGGGAGGAAATCCTCTC  
 5900  
 AlaIleProThrProArgAsnMetGln  
 LeuTyrArgProLeuGluThrCysAsnAsnSerCysTyrCysLysArgCysCysTyrHis  
 AGCTATACCGACCCCTAGAAACATGCAATAACTCATGCTATTGTAAGGGATGCTCCTACC  
 6000  
 MetAsnGluArgAlaAsp  
 CysGlnMetCysPheLeuAsnLysGlyLeuGlyIleCysTyrGluArgLysGlyArgArg  
 ATTGTCAGATGTGTTCTAAACAAAGGGCTGGGATATGTTATGCAACCAAAGGGCACAC  
 GluGluGlyLeuGlnArgLysLeuArgLeuIleArgLeuLeuHisGlnThrSerGluTyr  
 Met  
 ArgArgThrProLysLysThrLysThrHisProSerProThrProAspLys  
 GAACAAAGCACTCCAAAAGAAAAACTAAAGACTCATCCGTCCTACACCAGACAAGTGACTAT  
 6100  
 AspGluSerAlaAlaTyrCysHisPheIleSer  
 MetAsnGlnLeuLeuIleAlaIleLeuLeuAlaSerAlaCysLeuValTyrCysThrGln  
 GATGAATCAGCTGCTTATTGCCATTAACTAGCTAGTGCTGCTTAGTATATTGCACCCA  
 TyrValThrValPheTyrGlyValProThrTrpLysAsnAlaThrIleProLeuPheCys  
 ATATGTAACTGTTCTATGGCGTACCCACGTTGAAAAATGCAACCATCCCCCTCTTTC  
 6200  
 AlaThrArgAsnArgAspThrTrpGlyThrIleGlnCysLeuProAspAsnAspAspTyr  
 TGCAACCAGAAATAGGGATACTTGGGGAACCATACAGTGCTTGCCTGACAATGATGATTA  
 6300  
 GlnGluIleThrLeuAsnValThrGluAlaPheAspAlaTrpAsnAsnThrValThrGlu  
 TCAGGAAATAACTTGAATGTAACAGAGGCTTTGATGCCATGGAATAATACAGTAACAGA  
 GlnAlaIleGluAspValTrpHisLeuPheGluThrSerIleLysProCysValLysLeu  
 ACAAGCAATAAGAAGATGTCTGGCATCTATTGAGACATCAATAAAACCATGTGTCAAAC  
 6400  
 \*(fig.1A-sufte 5)

ThrProLeuCysValAlaMetLysCysSerSerThrGluSerSerThrGlyAsnAsnThr  
 AACACCTTATGTCTAGCAATCAAATGCCACAGAGACCACAGGGAAACAC  
 .  
 ThrSerLysSerThrSerThrThrThrProThrAspGlnGluGluGluIleSer  
 AACCTAAAGAGCACACCAACCACACCCACAGACCAGCAAGAGATAAC  
 6500  
 GluAspThrProCysAlaArgAlaAspAsnCysSerGlyLeuGlyGluGluGluThrIle  
 TGAGGATACTCCATGCGCACGCCAGACAATGCTCAGGATTGGGAGAGGAAGAACGAT  
 6600  
 AsnCysGlnPheAsnMetThrGlyLeuGluArgAspLysLysGlnTyrAsnGluThr  
 CAATTGCCACTTCATATCACAGGATTAGAAAGAGATAAGAAAAACAGTATAATGAAAC  
 .  
 TrpTyrSerLysAspValValCysGluThrAsnAsnSerThrAsnGlnThrGlnCysTyr  
 ATGCTACTCAAAGATGTGTTGTGAGACAAATAATGCCACAAATCAGACCCACTGTTA  
 6700  
 MetAsnHisCysAsnThrSerValIleThrGluSerCysAspLysHisTyrTrpAspAla  
 CATGAACCATTGCCAACACATCACTCATCACAGAACATGTGACAAGCAACTATTGGGATGC  
 .  
 IleArgPheArgTyrCysAlaProProGlyTyrAlaLeuLeuArgCysAsnAspThrAsn  
 TATAAGGTTTAGATACTGTGACCCACCGGTTATGCCCTATTAAAGATGTAATGATACCAA  
 6800  
 TyrSerGlyPheAlaProAsnCysSerLysValValAlaSerThrCysThrArgMetMet  
 TTATTCAGGCTTGACCCAACTGTTCTAAAGTAGTACGCTTCTACATGCCACCCAGGATGAT  
 6900  
 GluThrGlnThrSerThrTrpPheGlyPheAsnGlyThrArgAlaGluAsnArgThrTyr  
 GGAAACGCAAACCTCCACATGGTTGGCTTAATGCCACTAGAGCAGAGAATAGAACATA  
 .  
 IleTyrTrpHisGlyArgAspAsnArgThrIleIleSerLeuAsnLysTyrTyrAsnLeu  
 TATCTATTGGCATGGCAGAGATAATAGAACTATCATCAGCTTAAACAAATATTATAATCT  
 7000  
 SerLeuHisCysLysArgProGlyAsnLysThrValLysGlnIleMetLeuMetSerGly  
 CAGTTGCATTGTAAGAGGCCAGGGAAATAACACAGTGAACAAATAATGCTTATGTCAGG  
 .  
 HisValPheHisSerHisTyrGlnProIleAsnLysArgProArgGlnAlaTrpCysTrp  
 ACATGTGTTCACTCCCACTACCCAGGCCATCAATAAAAGACCCAGACAAGCATGGTGCCTG  
 7100  
 PheLysGlyLysTrpLysAspAlaMetGlnGluValLysGluThrLeuAlaLysHisPro  
 GTTCAAAGGCAAATGGAAAGACGCCATGCAGGAGGTGAAGGAAACCTTGCAAAACATCC  
 7200  
 ArgTyrArgGlyThrAsnAspThrArgAsnIleSerPheAlaAlaProGlyLysGlySer  
 CAGGTATAGAGGAACCAATGACACAAGGAATTAGCTTGCAAGGCCAGGAAAGGCTC  
 .  
 AspProGluValAlaTyrMetTrpThrAsnCysArgGlyGluPheLeuTyrCysAsnMet  
 ACACCCAGAAGTAGCATACATGTGGACTAACTGCAGAGGAGTTCTGACTGCAACAT  
 7300  
 ThrTrpPheLeuAsnTrpIleGluAsnLysThrHisArgAsnTyrAlaProCysHisIle  
 GACTTGTTCTCAATTGGATAGAGAATAAGACACACCCAAATTATGCCACCGTGCCATAT  
 .  
 LysGlnIleIleAsnThrTrpHisLysValGlyArgAsnValTyrLeuProProArgGlu  
 AAAGCAAATAATTACACATGCCATAAGGTAGGGAGAAATGTATATTGCCCTCCCAGGGA  
 7400  
 GlyGluLeuSerCysAsnSerThrValThrSerIleIleAlaAsnIleAspTrpGlnAsn  
 AGGGAGCTGCTGCAACTCAACAGTAACCAGCATATTGCTAACATTGACTGGCAAAA  
 7500  
 AsnAsnGluThrAsnIleThrPheSerAlaGluValAlaGluLeuTyrArgLeuGluLeu  
 CAATAATCAGACAAACATTACCTTACTGCCAGAGGTGGCAGAACTATAACAGATTGGAGTT  
 .  
 GlyAspTyrLysLeuValGluIleThrProIleGlyPheAlaProThrLysGluLysArg  
 GGGAGATTAAATGGTAGAAATAACCCAATTGGCTCCGACCTACAAAACAAAAAG  
 7600  
 (fig.1A-suite 6)

TyrSerSerAlaHisGlyArgHisThrArgGlyValPheValLeuGlyPheLeuGlyPhe  
 ATACTCCCTCTGCTCACGGAGACATACAACAGGTGTGTCGCTAGGGTTCTGGGTTT  
 .  
 LeuAlaThrAlaGlySerAlaMetGlyAlaAlaSerLeuThrValSerAlaGlnSerArg  
 TCTCGAACAGCAGGTTCTGCAATGGGCCGGCTCCCTGACCGTGTGGCTCACTCCCC  
 7700  
 .  
 ThrLeuLeuAlaGlyIleValGlnGlnGlnGlnLeuLeuAspValValLysArgGln  
 GACTTTACTGGCCGGATAGTGCAGAACAGCAACAGCTGTGGACGTGGTCAAGAGACA  
 7800  
 .  
 GlnGluLeuLeuArgLeuThrValTrpGlyThrLysAsnLeuGlnAlaArgValThrAla  
 ACAAGAACTGTTGGACTGACCCTCTGGGAACCAAAACCTCCACGGCAAGACTCACTGCC  
 .  
 IleGluLysTyrLeuGlnAspGlnAlaArgLeuAsnSerTrpGlyCysAlaPheArgGln  
 TATAGAGAACTACCTACAGGACCAGGCGGGCTAAATTCAATGGGATGTGGCTTAGACA  
 7900  
 .  
 ValCysHisThrThrValProTrpValAsnAspSerLeuAlaProAspTrpAspAsuMet  
 AGTCTGCCACACTACTGTACCATGGGTTAATGATTCTAGCACCTCACTGGGACAATAT  
 .  
 ThrTrpGlnGluTrpGluLysGlnValArgTyrLeuGluAlaAsnIleSerLysSerLeu  
 GACCTGGCAGGAATGGAAAAAAACAAGTCCGGTACCTGGAGGCAAATATCAGTAAAAGTTT  
 8000  
 .  
 GluGlnAlaGlnIleGlnGlnGluLysAsnMetTyrGluLeuGlnLysLeuAsnSerTrp  
 AGAACAGGCACAAATTCAAGCAAGAGAAAAATAIGTATGAACTACAAAATTAAATAGCTG  
 .  
 8100  
 AspIlePheGlyAsnTrpPheAspLeuThrSerTrpValLysTyrIleGlnTyrGlyVal  
 GGATATTTGGCAATTGGTTGACTTAACCTCCTGGGTCAAGTATATTCAATATGGAGT  
 .  
 LeuIleIleValAlaValIleAlaLeuArgIleValIleTyrValValGlnMetLeuSer  
 Val  
 GCTTATAATACTACCACTAAAGAATAGCTTTAAGAATAGTGATATATGTAGTACAAATGTTAAG  
 8200  
 .  
 AlaCysPheLeuPheProArgLeuTyrProThrAsp  
 ArgLeuArgLysGlyTyrArgProValPheSerSerProProGlyTyrIleGlnGlnIle  
 GlyLeuGluArgAlaIleGlyLeuPheSerLeuProProProValIleSerAsnArgSer  
 TAGGCTTAGAAAGGGCTATAGGCTGTTCTCTCCCCCCCCGGTTATATCCAACAGAT  
 .  
 ProTyrProGlnGlyProGlyThrAlaSerGlnArgArgAsnArgArgArgTrpLys  
 HisIleHisLysAspArgGlyGlnProAlaAsnGluGluThrGluGluAspGlyGlySer  
 IleSerThrArgThrGlyAspSerGlnProThrLysGlnLysLysThrValGluAla  
 CCATATCCACAAGGACCGGGGACAGCCAGCCAAACGAAGAACAGACGGTGGAAAG  
 8300  
 .  
 GlnArgTrpArgGlnIleLeuAlaLeuAlaAspSerIleTyrThrPheProAspProPro  
 AsnGlyGlyAspArgTyrTrpProTrpProIleAlaTyrIleHisPheLeuIleArgGln  
 ThrValGluThrAspThrGlyProGlyArg  
 CAACGGCTGGAGACAGATACTGGCCCTGGCCGATAGCATATATACATTCTGATCCGCCA  
 8400  
 .  
 AlaAspSerProLeuAspGlnThrIleGlnHisLeuGlnGlyLeuThrIleGlnGluLeu  
 LeuIleArgLeuLeuThrArgLeuTyrSerIleCysArgAspLeuLeuSerArgSerPhe  
 GCTGATTGGCTCTTGACCACATACAGCATCTGCAGGGACTTACTATCCAGGAGCTT  
 .  
 ProAspProProThrHisLeuProGluSerGlnArgLeuAlaGluThr  
 LeuThrLeuGlnLeuIleTyrGlnAsnLeuArgAspTrpLeuArgLeuArgThrAlaPhe  
 CCTGACCTCCAACTCATCTACAGAACTCAGAGACTGGCTGAGACTTACAACAGCCTT  
 8500  
 .  
 LeuGlnTyrGlyCysGluTrpIleGlnGlnAlaPheGlnAlaAlaAlaArgAlaThrArg  
 MetGlyAlaSerGlySerLysLysHisSerArgProProArgGlyLeuGlnGlu  
 CTTGCAATATGGGTCCGACTGGATCCAAGAACGATTCCAGGGCCGGGGCTACAAG  
 (fig.1A-suite 7)

GluThrLeuAlaGlyAlaCysArgGlyLeuTrpArgValLeuGluArgIleGlyArgGly  
 ArgLeuLeuArgAlaArgAlaGlyAlaCysGlyGlyTyrTrpAsnGluSerGlyGlyGlu  
 AGAGACTCTGGGGCGCTGCCACGGGCTTGTGGAGGGTATTGGAACGAATCGGGAGGG  
 8600  
 IleLeuAlaValProArgArgIleArgGinGlyAlaGluIleAlaLeuLeu  
 TyrSerArgPheGlnGluGlySerAspArgGluGlnLysSerProSerCysGluGlyArg  
 AATACTCGGGTTCCAAGAACGATCAGACAGGGAGCACAAATGCCCTCGTGTAGGGAC  
 8700  
 GlnTyrGlnGlnGlyAspPheMetAsnThrProTrpLysAspProAlaAlaGluArgGlu  
 GGCAGTATCAGCAGGGAGACTTTATGAATACTCCATGGAAGGACCCAGCAGCACAAAGGG  
 LysAsnLeuTyrArgGlnGlnAsnMetAspAspValAspSerAspAspAspAspGlnVal  
 AGAAAAAATTGTACAGCAACAAAATATGGATGATGTACATGATGATGATGACCAAC  
 8800  
 ArgValSerValThrProLysValProLeuArgProMetThrHisArgLeuAlaIleAsp  
 TAAGACTTTCTGTACACCAAAAGTACCACTAACGACATGACATAGATTGCAATAG  
 MetSerHisLeuIleLysThrArgGlyGlyLeuGluGlyMetPheTyrSerGluArgArg  
 ATATGTACACATTAATAAAACAAGGGGGACTGGAAGGGATGTTACAGTGAAAGAA  
 8900  
 HisLysIleLeuAsnIleTyrLeuGluLysGluGluGlyIleIleAlaAspTrpGlnAsn  
 GACATAAAATCTTAAATATACCTTAGAAAAGGAAGAAGGGATAATTGCAGATTGGCAGA  
 9000  
 TyrThrHisGlyProGlyValArgTyrProMetPhePheGlyTrpLeuTrpLysLeuVal  
 ACTACACTCATGGGCCAGGAGTAAGATAACCAATGTTCTTGGGTGGCTATGGAAGCTAG  
 ProValAspValProGlnGluGlyGluAspThrGluThrHisCysLeuValHisProAla  
 TACCAGTAGATGTCCCACAAGAAGGGGAGGACACTGAGACTCACTGCTTAGTACATCCAG  
 9100  
 GlnThrSerLysPheAspAspProHisGlyGluThrLeuValTrpGluPheAspProLeu  
 CACAAACAGCAAGTTGATGACCCGATGGGAGACACTAGTCTGGAGTTGATCCCT  
 LeuAlaTyrSerTyrGluAlaPheIleArgTyrProGluGluPheGlyHisLysSerGly  
 TGCTGGCTTATAGTTACCGAGGCTTTATTGGTACCCAGAGGAATTGGGCACAAAGTCAG  
 9200  
 LeuProGluGluTrpLysAlaArgLeuLysAlaArgGlyIleProPheSer  
 GCCTGCCAGAGGAGCTGGAGGGCAGACTGAAAGCAACAGGAATACCAATTAGTTAAA  
 9300  
 GACAGGAACAGCTATACTGGTCAGGGAGGAAGTAACAAACAGAAACAGCTGAGACTGC  
 AGGGACTTCCAGAAGGGCTGTAACCAAGGGAGGGACATGGGAGGAGCTGGTGGGAAC  
 9400  
 GCCCTCATATTCTCTGTATAAAATACCCCTAGCTTGCATTGTACTTCGGTCGCTCTGC  
 CGAGAGGCTGGCAGATGAGCCCTGGGAGGTCTCTCCACGAGCTACGAGGTAGAGCCTGG  
 9500  
 GTGTTCCCTGCTAGACTCTCACCAAGCACTTGGCCGGTGTCCCCAGACGGCCCCACGCTT  
 9600  
 CCTTGCTTAAACCTCCCTAAATAAGCTGCCAGTTAGAAGCA

(fig.1A-suite 8)

FIG 1B

ACTCGCTCTGGGAGAGGCTGCCAGATTGAGCCCTGGAGGTCTCTCCAGCACTAGCAG  
 CTAGAGCCCTGGTGTCCCTGCCTAGACTCTACCCAGCACTGGCCGGTGCTGGCAGACT  
 100  
 CGCTCCACGCTTGCTYCTTAAGACCTCTCAATAAAGCTGCCATTTAGAAGTAAGCTA  
 GTGTGTGTCCCATCTCTCTAGTCGCCCTGGTCAACTCGGTACTCGGTAAATAAAAAG  
 200  
 ACCCTGGTCTGTTAGGACCCCTGGCTGTAGGACCCCTTCGCTTGGGAAACCGAAGCA  
 300  
 CGAAAAATCCCTACCGAGATTGGCCGCCAACACGGACTTGAAGGAGAGTGAGAGACTCCTG  
 AGTACGGCTGAGTGAAGGCACTAAGGGGCCAGGAACCAACCACGACGGAGTGCTCTAG  
 AAAGGCGGGGTGGTACCGAGACGGCGTGAGGAGGGGAGAGAACAGGCCCTGGTTG  
 CAGGTAAGTCAACACAAAAAGAAATACCTGTCTTTATCCAGGAAGGGATAATAAGAT  
 500  
 GAGDMETGLYALAAARGASNSERVALLEUSERGLYLYSALAAASPGLULEUGLU  
 AGACTGGGAGATGGCGCAGAAACTCCGCTTGTCAGGGAAAGAACAGATGAATTAGA  
 LYSILEARGLEUARGPROGLYGLYLYSLYSTYRMETLEULYSHISVALVALTRPALA  
 600  
 AAAAATTAGACTACGACCCGGCCGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGC  
 ALAASNGLULEUASPARGPHEGLYLEUALAGLUSERLEULEUGLUASNLYSGLUGLYCYS  
 AGCAAATGAATTAGATAGATTGGATTAGCAGAAAGCCCTGGAGAACAAAGAACAGATG  
 700  
 GLNLYSILELEUSERVALLEUALAPROLEUVALPROTHRGLYSERGLUASNLLEULYSSER  
 TCAAAAATACTTTCGGCTTAGCTCCATTAGTGCACAGGCTCACAAAATTAAAAAG  
 LEUTYRASNTHRYVALCYSVALILETRPCYSILEHISALAGLUGLULYSVALYSHISTRH  
 CCTTATAATACTGTCTGGCTCATCTGGTGCACTACCCAGAACAGAACAC  
 800  
 GLUGLUALALYSGLNILEVALGLNARGHISLEUVALMETGLUTHRLYTHRALAGLUTHR  
 TGAGGAAGCAAAACAGATAGTGCAGAGACACCTAGTGTGAAACAGGAACAGCAGAAC  
 900  
 METPROLYSTHRSERARGPROTHRALARPROPHE SERGLYARGGLYLYASNTYRPROVAL  
 TATGCCAAAACAAGTAGACCAACAGCACCATTAGCGGCAGAGGGAAATTACCCAGT  
 GLNGLNILEGLYGLYASNTYRTHRHSLEUPROLEUSERPRODARGTHRLEUASHNALATRP  
 ACAACAAATAGGGTAACTATACCCACCTACCTAAAGCCCAGAACATTAATGCC  
 1000  
 VALLYSLIEUGLUGLULYSLYSPHEGLYALAGLUVALVALSERGLYFHEGLNLALEU  
 GGTAAAAATTAAATAGAGGAGAAGAAATTGGAGCAGAAGTAGTGTCAAGGATTTCAGGCACT  
 SERGLUGLYCYSLEUPROTYRASPILEASNGLNMETLEUASNCYSVALGLYASPHISGLN  
 GTCACAAGGCTGCCCTCCCTATGACATTAATCAGATGTTAAATTGTGTGGAGACCATCA  
 1100  
 ALAALAMETGLNILEILEARGASPILEILEASNGLUGLUALAALAAASPTRPASPLEUGLN  
 AGCGGCATGCAGATCATCAGAGATATTATAATCAGGAGGGCTGCAGATTGGCACTTGCA  
 1200  
 HISPRDGLNGLNLAAPROGLNGLNGLYGLNLEUARGGLUPROSERGLYSERASPILEALA  
 GCACCCACAACAAGCTCCACAACAAGGACACGCTTAGGGAGCCGTCAAGGATCAGATATTGC  
 GLYTHRTHRHSERTHRYVALGLUGLUGLNLILEGLNTRPMETTYRARGGLNGLNASHPROILE  
 AGGAACAACAGTACAGTAGAAGAACAAATCCAGTGGATGTACAGACACAGAACAGAACCCAT  
 1300

FIG. 1B-

PROVAL GLY ASN ILE TYR ARG ARG CTR PILE GLN LEU GLY LEU GLN HLY SCY SVAL ARG MET  
 ACC ACT AGC CAAC AT TT ACAGG AGAT GG AT CC AAC TGG GGT CC AAA AT GT GT CAGA AT  
 TYRA SN PRO THRA SN ILE LEU ASP VAL L YS GLN GLY PRO LY SG LU P ROP HE GL N SERT YR  
 CT AT A CCC A A C A A C A T T C T A G A T CT AAA A C A A G G G C C A A A G A G C C A T T C A C A G C T A  
 1400  
 VAL A SP ARG PHET Y RLY S SER LEU ARG ALA GLU GLN THRA S P R O A L A V A L Y S A S N T R P  
 TCT AGA CAG G T T C A C A A A G T T A A G C C A G A A C A A C A G A T C C A C C A G T A A A G A A T T G  
 1500  
 MET THRG L N THR LEU LEU GLN ASH ALA A S N P R O A S P C Y S L Y S L E U V A L L E U L Y S G L Y  
 CAT GACT CAA A C A C T G C T G A T T C A A A T G C T A A C C C A G A T T G C A A G C T A G T G C T G A A G G G  
 LEU GLY THRA SN PRO THR LEU GLU GLU MET LEU THRA L A C Y S G L N G L Y V A L G L Y G L Y P R O  
 G C T G G G T A C C A A T C C C A C C C T A G A A G A A T G C T G A C G G C C T G T C A A G G A G T A G C C G G G C C  
 1600  
 GLY GLN L Y S A A R G L E U M E T A L A G L U A L A E U L Y S G L U A L A E U A L A P R O A L A P R O I L E  
 A G G C A G A A G G C T A G A T T A A T G G C A G A A G C C C T G A A A G A G G C C T C G C A C C A G G C C A A T  
 POL VAL L E U G L U L E U T R P  
 PRO PHE A L A A A L A G L N L N L Y S G L Y P R O A R G L Y S P R D I L E L Y S C Y S T R P A S N C Y S G L Y  
 C C C T T T G C A G C A G C C A A C A G A A G G G A C C A A G A A G C C A A T T A A G T G T T G G A A T T G T G G  
 1700  
 GLU GLY ARG T H R L E U C Y S L A M E T G L N S E R P R D L Y S T H R G L Y M E T L E U G L U M E T  
 L Y S G L U G L Y H I S S E R A L A A R G G L N C Y S A R G A L A P R O A R G A R G G L N G L Y C Y S T R P L Y S C Y S  
 G A A G G A A G G A C A C T C T G C A A G G C A A T G C A G A G C C C A A G A A G A C A G O G A T G C T G G A A T G  
 1800  
 T R P L Y S A S N G L Y P R O C Y S T Y R G L Y G L N H E T P R O L Y S G L N T H R G L Y G L Y P H E P H A E R G P R O  
 G L Y L Y M E T A S P H I S Y A L M E T A L A Y S C Y S P R O A S H A R G G L N A L A G L Y P H E L E U G L Y L E U  
 T G G A A A A T G G A C C A T G T A T G C C A A A T G C C C A A C A G A C A G C C G G G T T T T A G G C C T  
 T R P P R O L E U G L Y L Y S G L U A L A P R O G L N P H E P R O H I S G L Y S E R S E R A L A S E R G L Y A A S P  
 G L Y P R O T R P G L Y L Y S L Y S P R O A R G A S N P H E P R O M E T A L A G L N V A L H I S G L N G L Y L E U T H R  
 T G G C C T T G G G A A A G A A G C C C C C A A T T T C C C A T G G C T C A A G T G C A T C A G G G G C T G A C  
 1900  
 A L A A S N C Y S S E R P R O A R G A R G T H R S E R C Y S G L Y S E R A L A L Y S G L U L E U H I S A L A L E U G L Y  
 P R O T H R A L A P R O Q U G L U P R O A L A V A L A S P L E U L Y S A N T Y R M E T H I S L E U G L Y  
 G C C A A C T G C T C C C C C A A G A A G C C C A G G C T G C G A T C T G C T A A A G A A C T A C A T G C A C T T G G  
 G L N A A L A A L A G L U A R G L Y S G L N A R G G L U A L A E U G L N G L Y G L Y A S P A R G G L Y P H E A L A A L A  
 L Y S G L N G L N A R G G L U S E R A R G G L Y L Y S P R O T Y R L Y S G L U V A L T H R G L U A S P L E U H I S  
 C A A G C A C G A G A G A A G C A G A G G G A A G C C T T A C A A G G A G G T G A C A G G G A T T T G C T G C A  
 2000  
 Z 0 0  
 P R O G L N P H E S E R L E U T R P A R C G P R O V A L V A L T H R A L A H I S I L E G L U G L Y G L N P R O V A L  
 L E U A N S E R L E U P H E G L Y G L Y A S P G L N  
 C C T C A A T T C T C T T G G A G G A G C C A G T A G T C A C T G C T C A T A T T G A A G G A C A G C C T G T A  
 2100  
 G L U V A L L E U A S P T H R G L Y A L A A S P A S P S E R I L E V A L T H R G L Y I L E G L U L E U G L Y P R O  
 G A A G T A T T A G A T A C A G G G G C T G A T G C T A T T G T A A C A G G A A T A G A G T T A G G T C C A  
 H I S T Y R T H R P R O L Y S I L E V A L G L Y G L Y I L E G L Y G L Y P H E I L E A S N T H R L Y S G L U T Y R L Y S  
 C A T T A A C C C C A A A A T A G T A C G A G G A A T A G G G A G T T T A T T A A T C T A A G A A T A C A A A  
 2200  
 A S N V A L G L U I L E G L U V A L L E U G L Y L Y S A R G I L E L Y S G L Y T H R I L E M E T T H R G L Y A S P T H R  
 A A T G T A G A A A T A G A A G T T T A G G C C A A A G G A T T A A A G G C A A T C A T G A C A G G G G A C A C C  
 P R O T E A S N I L E P H E G L Y A R G A S N L E U L E U T H R A L A L E U G L Y M E T S E R L E U A S N L E U P R O  
 C C G A T T A A C A T T T G G T A G A A A T T A C T A A C A G C T C T G G G A T G T C T C A A A T C T C C C  
 2300  
 T I L E A L A L Y S V A L G L U P R O V A L L Y S S E R P R O L E U L Y S P R O G L Y L Y S A S P G L Y P R O L Y S L E U  
 A T A G C T A A G G T A G A G G C C T G T A A A G T C G C C C T T A A G C C A G G A A A G G A T G G A C C A A A T T G  
 2400  
 L Y S C L N T R P P R O L E U S E R L Y S G L U L Y S I L E V A L A L A E U A R G G L U I L E C Y S G L U L Y S N E T  
 A A G C A G T G C C C A T T A T C A A A G A A A A G A T A G T T G C A T T A A G A G A A A T C T G T G A A A A G A T G

(Fig. 12 - Suite 1)

GLULYSASPGLYGLNL EUGL UGL UAL APROPROTHRASNPROTYRASNTHRPROTHRHE  
 GAAAAAGATGGTCAGTTGGAGGAAGCTCCCCCACCACATCCATAACACCCCCACATT  
 2500  
 ALAILELYSLYSLYSASPLYSASNLYSTRPARGHETLEUILEASPPHEARGGLULEUASN  
 CCTATAAACAAAAGGATAAAAACAAATGGAGAATGCTGATAGATTTAGGGAACTAAAT  
 ARGYALTHRGGLNASPPHE THRGLUVALCLNLEUGLYILEPROHISPROALAGLYLEUALA  
 AGGGTCACTCAAGACTTTACCGAAGTCCAATTAGGAATACCACACCTGCAGGACTAGCA  
 2600  
 LYSARGLYSARGILETHRVALLEUASPILEGLYASPALATYRPHESERILEPROLEUASP  
 AAAAGGAAAAGGATTACAGTACTGGATATAGGTGACGCCATTTCTCTATACCTCTAGAT  
 2700  
 GLUGLUPHEARGGLNTYRHLAPHETHRLEUPROSERVALASNALAGLUPROGLY  
 GAAGAATTAGGCAGTACACTGCCTTACTTACCATCACTAAATATGGAGAGCCAGGA  
 LYSARGTYRILETYRLYSVALLEUPROGLNGLYTRPLYSGLYSERPRDIALAILEPHEGLN  
 AAACGATACATTATAAGGTTCTGCCTCACGGATGGAAAGGGTCAACAGGCATCTTCAA  
 2800  
 TYRTHRMETARGHISVALLEUGLUPROPHEARGLYSALAASNPROASPVALTHRLEUVAL  
 TACACTATGAGACATGTGCTAGAACCCCTCAGGAAGGCAAATCCAGATGTGACCTTAGTC  
 GLNTYRME-TASPASPILEUILEALASERASPARGHRSPLUEGLUHISASPARGVAL  
 CAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGACAGGGTA  
 2900  
 VALLEUGLNLNEULYSGLULEUEA NSN SERILEGLYPHESER SERPROGLUGLULYSPHE  
 GTTTACAGTTAAAAGAACCTCTAAATAGCATAGGGTTCATCCCCAGAACAGAAATTG  
 3000  
 GLNLYSAPPORPROPHEGLNTRPMETGLTYRGLULEUTRPPROTHRLYSTRPLYSLEU  
 CAAAAGATCCCCATTCAATGGATGGGTACGAATTGTGGCCGACAATGGAAAGTTG  
 GLNLYSILEGLULEUPROGLNARGGLUTHRTRPHTHRVALASNASPILEGLNLYSLEUVAL  
 CAAAGATAGACTTCCCACAAAGAGACCTGGACAGTGAATGATATAAGAAGTTAGTA  
 3100  
 GLYVALLEUASNPALALAAGLNILETYRPROGLYILELYSTHRLYSHISLEUCYSARG  
 GGAGTATAAAATTGGCAGCTCAAATTATCCAGGTATAAAACCAACATCTCTGTAGG  
 LEUILEARGLYLYSMETTHRLEUTHRGLUGLUVALGLNTRPHTRGLUMETALAGLUALA  
 TTAATTAGGAAAATGACTCTAACAGAGGAAGTCTAGTGGACTGAGATGGCAGAACAGCA  
 3200  
 GLUTYRGLUGLUASHNLYSILEELELEUSERGLNGLUGLNGLUGLYCYSTYRTYRGLNGLU  
 GAATATGAGGAAAATAAAATAATTCTCAGTCAGGAACAAGAAGGATGTATTACCAAGAA  
 3300  
 SERLYSPROLEUGLUALATHRALYVALILELYSSERGLNAPASHNGLNTRPSERTYRLYSILE  
 AGCAACCCATTAGAACGCCACGGTATAAACAGACTCAGGACAATCACTGGCTTATAAAATT  
 HISGLNGLUASPLYSILEUlysVALGLYLYSPHEALALYSILELYSASNTHRHISTHR  
 CACCAAGAACAAAATACTGAAAGTAGGAAAATTGCAAAAGATAAGAACATACACATACC  
 3400  
 ASNGLYLYVALARGLEUUEUALAHISVALILEGLNLYSILEGLYLYSGLUALAILEVALILE  
 AATGGAGTTAGACTATTACCATCTAACAGAAAATAGGAAGGAAAGCAATAGTGTAC  
 TRPGLYGLNVALPROLYSPHEHISLEUPROYALGLULYSASPVALTRPGLUGLNRPTRP  
 TGGGCACAGCTCCAAAATTCCACTTACCACTTGAGAAGGATGTATGGAAACAGTGGTGG  
 3500  
 THRASPTYRTRPGLNVALTHRTRPILEPROGLUTRPASPPHEILESERTHRPROPROLEU  
 ACAGACTATTGGCAGGTAACTGGATACCGGAATGGGTTCATCTCAACACCCACATTAA  
 3600  
 VALARGLEUVALPHEASNLEUVALLYSASPPROILEGLUGLYGLUGLUTHRTYRTYRVAL  
 GTAAGATTAGTCTCAATCTAGTGAGGACCCCTATAGGGAGAAGAACCTATTATGTA  
 ASPGLYSERCYSSERLYSGLN SERLYSGLUGLYLYSALAGLYTYRIETHRASPARGGLY  
 GATGATCATGAGTAAACAGTCAAAAGAAGGAAAGCAGGATATACACAGACACGGGGC

(Fig. 1B - Seite 2)

3700

LYSASPLYSVALLYSVALLEUGLUGLNTHRTHRASNGLNGLNALAGLULEUGLUALAPHE  
AAAGACAAGTAAACTGTTAGAACAGACTACTAATCAACAAGCACAATTGCAAGCATT

LEUMETALALEUTHRASP SERGLYPROLYSALAASNILEILEVALASP SERGLNTHRVAL  
CTCATGGCATTGACAGACTCAGGCCAAAGCCAAATTATATAGTACACTCACAAATATGTT  
3800

METGLYILEILETHRGGLYSPROTHRGLUSERGLUSERGLEUVALASHGLNILEILE  
ATGGCAATAATAACAGGATCCCCTACAGAATCAGAGACGAGGCTAGTTAACCAAATAATA  
3900

GLUGLUMETILELYSLYSTHRLGLIETRYRVALALATRPVALPROALAHISLYSGLYILE  
GAAGAAATGATCAAAAAGACAGAAATTATGTGCATGGTACCAAGCACACAAAGGTATA

GLYGLYASNLNGLNLUILEASPHISLEUVALSERGLNGLYILEARGCLNVALLEUPHELEU  
CGAGGAAACCAAGAAATACACCACCTAGTTAGTCAGGGATTAGACAAGTTCTCTCTTG  
4000

GLULYSILEGLUPROALAGLNGLUGLHISERLYSTYRHISSEASNILELYSGLULEU  
GAAAGATAGAGCCAGCACAGAAACATAGTAATACCATAGTAACATAAAAGAATTG

VALPHELYSPHEGLYLEUPROARGLEUVALALALYSGLNILEVALASP THRCYSASPLYS  
GTATTCAAATTGGATTACCCAGACTAGTGCCTAACAGATAGTAGACACATGTGATAAA  
4100

CYS HIS GLN LYS GLY GLU ALA LEHIS GLY GLN VAL ALA NS ERAS PLE UGLY THR TRP GLN  
TGT CAT CAA AA AGG AGA AGC TAT AC AT GGC AGG TA ATT CAG AC CT A GG ACT TT G CAA  
4200

MET ASP CYST HR HIS LEU GLU GLY L Y S I LEU AL I LEU AL A L A V AL H I S V AL A L A S E R G L Y  
AT G C T T G T A C C C A T C T A G A G G G A A A A T A C T C A T A G T T G C A G T A C A T G T A G C T A G T G G A  
PHE I E GLU AL A GLU V AL I LE PRO GLN GLU TH R G L Y A R G G L N T H R A L A E U P H E L E U L E  
TTCATAGAACAGAACAGTAATTCCACAAAGAAACAGGAAGACAGACAGCACTATTCTGTTA  
4300

LYS LEU AL A SER AR GTR P PRO I LE TH R H I S LEU H I S T H R P A S N G L Y A L A A S N P H E A L A  
AAATTGGCAAGCAGATGCCCTATTACACATCTGCACACAGATAATGGTCTAACTTGCT

SERGLNGLUVALLYSMETVALALATRPTRPALAGLYILEGLUHISTHRPHEGLYVALPRO  
TCCGAAGAAGTAAAGATGGTGCATGGTGGCAGGGATAGAGCACACCTTTGGGTACCA  
4400

TYR ASN PRO GLN SER GLN GLY VAL VAL GLU ALA MET ASN H I S H I S I S L E U L Y S A S H G L N I L E  
TACAATCCACAGAGTCAGGGAGTAGTGGAAAGCAATGAATCACCACCTGAAAATCAAATA  
4500

ASP ARG ILE ARG GLU GLN ALA ASN SER V AL GLU TH R I L E V AL L E U M E T A L A V A L H I S C Y S  
GATAAACATCAGGGAAACAAGCAATTCACTAGAAACCATAGTATTATGGCAGTTCTATTGC

RET ASN PHE LY S A R G A R G G L Y G L Y I L E G L Y A S P H E T H R P D A L A G L U A R G L E U I L E A S N  
ATGAATTTAAAGAAGGGAGGAATAGGGATATGACTCCAGCAGAAAGATTAAAC  
4600

MET T I L E T H R T H R G L U G L N G L U I L E G L N P H E G L N G L N S E R L Y S P H E L Y S A S N  
ATGATCACTACAGAACAAAGAAATACAATTCAACAATCAAAAAACTCAAATTAAAAAT

PHE ARG VAL TYR TYR R A G G L U G L Y A R G A S P G L N L E U T R P L Y S G L Y P R O G L Y G L U L E U L E  
TTCCGGTCTATTACAGAACAGCAGACATCAGCTGTGCAAGGGACCCGGTGAGCTATTG  
4700

T R P L Y S G L Y G L U G L Y A L A V A L I L E L E U L Y S V A L G L Y T H R A S P I L E L Y S V A L V A L P R O A R G  
TGGAAAGGGGAAGGGACAGTCATCTAAAGGTAGGAACAGACATTAAGGTAGTACCCAGG  
4800

A R G L Y S A L A L Y S I L E I L E L Y S A S P T Y R G L Y G L Y L Y S G L U M E T A S P S E R S E R H I S  
Q M E T G L U G L U G L U L Y S A R G T R P I L E V A L V A L P R O T H R  
AGAAAGGCTAAAATTATCAAAGATTATGGAGGAGGAAAGAGATGGATAGTAGTTCCAC

MET GLU AS P T H R G L Y G L U A R G G L U A L A  
T R P A R G I L E P R O G L U A R G L E U G L U A R G T R P H I S S E R L E U I L E L Y S T Y R L E U L Y S T Y R L Y S  
ATGGAGGATACGGAGAGGCTAGAGAGGTGCCATAGCCTCATAAAATATTGAAATATAA  
4900

(fir. 1B - Seite 3)

THRLYSASPLEUGLNLYSALACYSTYRALPROHISIISLYSVALGLYTRPALATRPTRP  
 AACTAAAGATCTACAAAAGCTTGCTATGTGCCCATCATAAGTCGGATGGCATGGTC  
 THRCYSSERARGVALILEPHEPROLEUGLNGLUGLYSERHISLEUGLUVALGLNGLTYR  
 GACCTGCAGCAGACTAATCTTCCCACGAGGAACCCATTAGAAGTACAACGGTA  
 5000  
 TRPASNLEUTHRPRDGLUARGGLYTRPLEUSERTHRTHRYRALAVALARGILETHRPTYR  
 TTGGAATTGACACCAGAAAGAGGGTGCCTCAGTACTTTAGCACTGAGGATAACCTGGTA  
 5100  
 SERLYSASPPEPTRPTHRASPVALTHRPROGLUTYRALAASPITLEULEUHMISERTHR  
 CTCAAAGGACTTTGACAGATGTAACACCGAAATATGCACTTTACTGCATACAC  
 TYRPHEPROCYSPHETHRALAGLYGLUVALARGARGALAILEARGGLYGLUARGLEULEU  
 TTATTTCCCTTGCTTACAGCGGGAGAACTGAGAAGGGCCATCAGGGAGAACACTGCT  
 5200  
 SERCYSCYSARGPHEPROARGALAHISLYSHISGLNVALPROSERLEUGLNTYRLEUALA  
 GTCTTGCAGGTTCCCAGAGCTCATAAAGCACCEAGGTACCAAGTCTACAGTACTTAGC  
 LEUARGVALVALSERHISVALARGSERGLNGLYGLUASNPROTHRTRPLYSGLNTRPARG  
 X METSERASPPROARGGLUARGILEPROPROGLYASNNSERGLYGLU  
 ACTGAGAGTAGTAAAGTCATGTCAGATCCCAGGGAGAAATCCCACCTGGAAACAGTGGAG  
 5300  
 ARGASPASHARGSERLEUARGVALALALYSGLNASHSERARGGLYASPLYSGLNARG  
 GLUTHRILEGLYGLUALAPHEGLUTRPLEUASNARGTHRVALGLUGLUILEASHARGGLU  
 AAGAGACAATAGGAGAAGCCTTCAGTGCCTAAACAGAACAGTAGAGGGAGATAAACACAG  
 5400  
 GLYGLYLYSPROPROTHRGLUGLYALAASNPHPROGLYLEUALALYSVALLEUGLYILE  
 ALAVALASNHSILEUPROARCGGLEUILEPHEGLNVALTRPGLNARGSERTRPGLUTYR  
 AGGGCTAAACCACTACCGAGGGAGCTAATTTCAGGTTGGCAAAGGTCTGGGAAT  
 LEUALA  
 TRPHISASPGLUGLNGLYMETSERGLNSERTYRTHRLYSTYRARGTYRLEUCYSLEUILE  
 ACTGGCATGTAACAAGGGATGTCACAAGCTATACAAAATACAGATACTTGTGTTAA  
 5500  
 GLNLYSALALEUPHEMETHISCSYSLYSGLYCYSARGCYSLEUGLYGLUGLYHISGLY  
 TACAAAAGGCTTATTATGTCATTCCAAGAAAGGCTGTAGATGCTAGGGAAAGGACACG  
 ALAGLYGLYTRPARGPROGLYPROPRDPROPRDPROPROGLYLEUALA R METGLU  
 GGGCAGGGGGATGGAGACCAAGGACCTCCCTCCCTCCCCCTCCAGGACTAGCATAATGG  
 5600  
 GLUARGPROPROGLUASNGLUGLYPROGLNARGGLUROTRPASPGLUUTRPLYVALVALGLU  
 AAGAAAGACCTCCACAAAATGAAGGCCCACAAAGGGAAACCTGGGATGACTGGGTAGTGG  
 5700  
 VALLEULYSGLULEULYSGLUGLUALALEULYSHISPHEASPPROARGLEULEUTHRALA  
 AAGTCTGAAAGAACTGAAAGAAGCTTAAAGCATTGTCATCTCGGCTTCTAACCG  
 TAT1 METGLUTHRPRDLEUARGGLUGLNGLUANSER  
 LEUGLYASNHSILETYRASHRGHISGLYASPTHRLEUGLUGLYALAGLYGLULEUILE  
 CACTTGGTAATCATATCTATAATAGACATGGAGACACCCCTGGGAGCAGGAGAACTCA  
 5800  
 LEUGLUSERSERASNGLARGERSERTYRILESERGLUALAALALAILEPROGLU  
 ARGILELEUGLNARGALALEUPHEIHLHISPHARGEARGLYCYSERHISERARGILE  
 TTAGAATCTCCAACCGAGGGCTTCTCATACATTTCAGAACGGCTGCAGCCATTCCAGAA  
 SERALAASNLEUGLYGLUGLUILELEUSERGLNLEUTYRARGPROLEUGLUALACYSTYR  
 GLYGLNPROGLYGLYASNPROLEUSERTHRILEPROPROSERARGSERMETLEU  
 TCGGCGAACCTGGGGAGGAAATCCTCTCTCAACTATACCCCTCTAGAAGCAGTGTAT  
 5900  
 ASNTHRCYSTYRCYSLYSLYSCYSCYSTYRHISCSYSGLHPHECYSPHELEULYSLYSGLY  
 AACACATGCTATTGCAAAAAGTGTGCTACCATTTGCACTGTTCTAAAGGGC  
 LEUGLYILESERTYRGLULYSSERHISARGARGARGTHRPROLYSLYSALALYSALA  
 ARTIMETARGSERHISTHRLYGLUGLUGLULEUARGARGCARGLEUARGLEU

(fig.1B-suite 4)

1TGGGATAACTTATGACAAGTCACACAGGAGAAGAAACTCCGAGAACGGCTAAGGCT  
 ASNTHRSERSERALASERASNGLU  
 ILEHISLFLUEUMISGLNTHRSERLYSTYRGLYLEUSERTRPLYSSERALAAATYRARG  
 ENV MET GLYCYSLEUGLYASNGLNLEULEUILEALA  
 AATACATCTCTGCATCAACCGACTAAGTATGGCTCTGGAAATCAGCTGCTTATCG  
 . . . . .  
 6100  
 HISLEULEU  
 ILECYSSERLYSCYSLEUTRPILEILECYSILEGLNTYRVALTHRVALPHETYRGLYVAL  
 CCATCTGCTCTAAGTGTCTATGGATTATTGATTCAATATGTACAGCTTTATGGTC  
 . . . . .  
 PRODALATRPARSGNALATHRILEPROLEUPHECYSALATHRLYSASHARGASPTHRTRP  
 TACCACTGGAGGAATGGACAATTCCCCCTTCTGTGCAACCAAGAAAGGGACTT  
 6200  
 GLYTHRTHRLNCYSLEUPDASPASNAPASPTYRSERGLULEUALAEUASNVALTHR  
 GGGAAACAACCTCAGTGCCTACCCAGATAATGATGATTATTCAGAATTGCCCTTAATGTA  
 . . . . .  
 6300  
 GLUSERPHEASPALATRPGLUASNTHRALTHRGLUGLNLALILEGLUASPYALTRPGLN  
 CAGAAAGCTTGATGCTTGGAGAATAACAGTCACAGAACAGGCAATAGAGGACGTATGGC  
 . . . . .  
 LEUPHEGLUTHRSERILELYSPROCYSVALLYSLEUSERPROLEUCYSILETHRMEARG  
 AACTCTTGAGACCTCAATAAACCTTGTGTAATTATCCCCATTATGCATTACTATGA  
 . . . . .  
 6400  
 CYSASNYSSERGLUTHRASPLYSTRPGLYLEUTHRLYSERSERTHRTHRTHRALASER  
 GATGCAATAAAAGTCAAGACAGATAATGGGATTGACAAAATCATCAACAACACCAT  
 . . . . .  
 THRTHRTHRTHRTHRTHRALYSSERGLUTHRARGASPILEYALASNGLUTHRSER  
 CAACAACAAACAACACAGCAAATCAGTAGACACAGACAGACATAGTCATCAGACTA  
 6500  
 6600  
 PROCYSVALVALHISASPASNCYSTHRLYGLYLEUGLUGLNLUPROMETILESERCYSLYS  
 GTCCCTTGTTAGTTCTATGATAATTGCACAGCTTGGCAACAAGGCCAATGATAACCTGTA  
 PHEASNMETTHRGLYLYEULYSARGASPLYSLYSLYSLYRASNLUTHRTRPTYRSER  
 ATTCAACATGACAGGGTTAAAAGAGACAGAAAGAAAAAGGACTAACATCAAACCTGGCTACT  
 . . . . .  
 6700  
 ALAASPLEUVALCYSGLUGLNLGYASNSERTHRLYASNGLUSERARGCYSTYRMETASH  
 CTGCAGATCTGGTTGTGAAACAAGGAATAGCACTGGTAATGAAAGTAGATGTTACATGA  
 . . . . .  
 HIS CYSASNTHRSERYALILEGLNLGLUCYSCYSASPLYSRTRPASPALAILEARG  
 ATCACTGTAATACTCTGTTATCCAAGAGTGTTGACAAAAGATTATGGGATGCTATTA  
 . . . . .  
 CYSARGTYRCYCALAPROPROGLYTYRALALEULEUARGCYSASNSPTHRASNTYRSER  
 GATGAGATATTGCACTTCCAGTTATGCTTCTGTTAGATGTAATGACACAAATTATT  
 6800  
 GLYPHENETPRODASNCYSERLYSVALVALSERSERCYSYTHRARGMETGLUTHR  
 CAGGCTTATGCCCTAACTGTTCAAGGTAGTGGCTCTCATGCACAAGGATGATGGAGA  
 . . . . .  
 6900  
 GLNTHRSERTHRTRRPPHEARGPHEASNLYTHRARGALAGLUASNARGTHRTHRYLETYR  
 CACAGACTTCTACTGGTTGGTTAATGGAACTAGACAGCAAAATGACACCTATATT  
 . . . . .  
 TRPHISGLYARGASPAHARGTHRILEILESERLEUASNLYSHIYSTYRASNLEUTHRNET  
 ACTGGCATGGTAGAGATAATAGGACTATAATTGCTAAATAAGCATTATAATCTAACAA  
 . . . . .  
 7000  
 LYS CYSARGARGPROGLYASHLYSTHRLVALLEUROVALTHRLEMETSERALALEUVAL  
 TGAATGTAAGAGACCAGGAATAAGACAGTTTACCACTCACCATTATGCTGCATTGG  
 . . . . .  
 PHEHISERGLNPROVALASNGLUARGPROLYSGLNLAATRCPYSARGPHEGLYGLYASH  
 TTTTCCACTCACAAACCACTCAATGAGAGGCCAAAGCAGCCATGGCTGAGCTTGGAGGA  
 . . . . .  
 7100  
 TRPLYSGLUALAILELYSGLUVALLYSGLNTHRILEVALLYSHIISPRDARGTYRTHRLY  
 ATTGGAGGAGGCAATAAAAGAGGTGAAGCAGACCCATGTCACATCCCAGGTATACTG  
 . . . . .  
 7200  
 THRASNASHTRASPLYSILEASHLEUTHRALAPROARGGLYGLYASPROGLUVALTHR  
 GAACTAACAAACTGATAAAAATCAATTGACGGCTCTAGAGGAGGAGATCCGGAAAGTTA

(fig.1B-suite 5)

PHE<sup>1</sup>ETTRPTHRASNCYSARGGLYGLUPHELEUTYRCYSLYSME TASNTRP PHELEU ASN  
 CCTTCATGAGCAAAATTGCGAGGAGAGTTCTCTACTCTAAAATGAATGGTTCTAA  
 7300  
 TRPVALGLUASPARGSERLEUTHRTHRLNLYS PROLYS GLU ARGHISLYSARGASNYR  
 ATTGGCTAGAACAGATAGGAGCTA AACTACCCAGAACGCCAAAGGAACGGCATAAAAGGAATT  
 VALPROCYSHISILEARGGLNILEILEASNTHRTRPHISLYSVALGLYLYSASNVATYR  
 ACGTACCATGTCATATTAGACAATAATCACACTTGGCATAAAGTAGGCAAAATGTT  
 7400  
 LEUPRODPROARGGLUGLYASPLEUTHRCYSASN SERTHRVALTHR SERLEUTILEALA ASN  
 ATTTGCCCTCAAGAGAGGGACCTCACGTGTA CTCCACACTGACCAGCTCATAGCAA  
 7500  
 ILEASNTRPTHRASPGLYASNGLNTHR SERILETHR METSERALAGLUVALALAGLULEU  
 ACATAAAATTGGACTGATGGAAACCAACTAGTATCACCATGAGTCCAGAGGGTGGCAGAAC  
 TYRARGLEUGLULEUGLYASPTYRLYSL EUVALGLUILETHR PROJILEGLYLEUALAPRO  
 TGTATCGATTGAAATTGGAGATTATAAAATTAGTAGAAATCACTCCAATTGGCTGGCC  
 7600  
 THRASNVALYSARGTYRTHRTHRLGYGLYTHRSERARGASNLYSARGGLYVALPHEVAL  
 CCACAAATGTGAAGAGGTACACTADTGGTGGCACCTCAAGAAATAAAAGAGGGCTTTC  
 LEUGLYPHELEUGLYPHELEUALATHRALAGLYSERALAMETGLYALAA LASERLEUTHR  
 TGCTACGGTTCTGGGTTCTCGCAACGGCAGGTTCTGCAATGGCGGGCGTCGTTGA  
 7700  
 VALTHR ALAGLN RARGTHR L EULEUALAGLYILEVALGLNGLNGLNGLNLEUEU  
 CGTGACCGCTCAGTCCC GGACTTTATTGGCTGGGATAGTGCAGCAACAGCAACAGCTG  
 7800  
 ASPVALVALLYSARGGLNGLNGLULEUARGLEUTHRVALTRPGLYTHR LYSASNLEU  
 TCGACGTTGGTCAAGAGACAACAAGAATTGTTGGCACTGACCGCTGGGACAAAGAAC  
 GLNTHRARGVALSERAL ILEGLULYSTYRLEULYSASPGLNALAGLNLEU ASNALATR  
 TCCAGACTAGGGTCTCTGCCATCGAGAAGTACTAAAGGACCAGGGCAGCTAAATGCT  
 7900  
 GLYCYSALAPHEARGGLNVALCYSHISTRHRYALPROTRP PROASNALASERLEUTHR  
 GGGGATGTCGTTTAGACAAGTCTGTCACACTACTGTACCATGCCAAATGCAAGTCTAA  
 PROASP TRPASN ASNLUTHRTRPGLNGLUTRPGLUARGLYSVALASP PHELEUGLUA  
 CACCAAGATTGGACAATGAGACTTGGCAAGAGGGCAAGAGTGGGAGCGGAAGGTTGACTCTTGGAGG  
 8000  
 ASNILETHR ALALEU EUGLUGLUALAGLNILEGLNGLNGLULYSASN METTYRGLULEU  
 CAAATAAACGGCCCTCTAGAACAGAGGCAAAATTCAACAAGAGAACATGTATGAAT  
 8100  
 GLNLYSLEU ASN SERTRPASPVALPHEGLYASNTRP PHEASPLEUTHR SERTRP ILELYS  
 TACAAAAGTTGAATAGCTGGATGTGTTGCCATTGGTTGACCTTACTCTGGATAA  
 TYRILEGLNTYRGLYILETYRILEILEVALGLYVAL ILEULEUARGILEVALILETYR  
 ACTATATAACATGGAAATTATATAATTG TAGGAGTAATCTGTTAAGAATAGTGTACT  
 8200  
 ILEVALGLN METLEU ALA ARGLEUARGGLNGLYTYRARGPROVALPHE SERSER PROPRO  
 ATATAGTACAATGCTAGCTACGTTAACAGGGCTATAGCCAGTGTCTCTCCCCAC  
 TAT2ARGPROILEPROASNRGILEARGLEUCYSGLNPROLYS LYSAL  
 ART2VALASPPROTYRPRDTHRLGLYSERGLYSERALASNGLNARGARGGLN  
 SERTYRPEGLN\*\*\*THR HISTHRLGLNGLNASPPROALALEUPROTHRLYSGLUGLYL  
 CCTCTTATTCCAGTAGACCCATACCAACAGGATCCGGCTGCCAACCAAGAACAGGCA  
 8300  
 LYSLYSGLU THRVALGLU ALAVALALATHRALA PROGLYLEUGLYARG (TAT(f1))  
 LYSARGARGAR GRTPAR GGLN ARGP RLGLN LIEULEU ALA SPARGILE TYR  
 LYSGLYASPGLYGLYGLYSERGLYGLYASNSER SERTRP PROTRPGLNILEGLUTYRILE  
 AAAAGGAGACGGTGGAGGCAGCGTGGCAACAGCTCTGGCTGGCAGATAGAATA  
 8400

SERPHPROASPPROPROTHRASPTHRPROLEUA<sup>1</sup>PLEUALAS<sup>2</sup>LEGLNLNL<sup>3</sup>UGLN<sup>4</sup>NAS  
HISPHELEUILEARGLNL<sup>5</sup>LUILEARGLEULEU<sup>6</sup>THTRP<sup>7</sup>PLEUPHESERASNCYS<sup>8</sup>SARGTHR  
TTCATTCTCTGATCCGCCACTGATACGCCCTCTGACTTCGCTATTCA<sup>9</sup>CAACTGCAGAA  
LEUALAILEGLUSERTILEPROASPPROPROTHRASHILEPROGLUALALEUCYSASPLFU  
LEULEUSERARGALATYRGLNL<sup>10</sup>LEU<sup>11</sup>UGLN<sup>12</sup>PROILEPHEGLN<sup>13</sup>ARGLEUSERALATHRTYR  
CCTTGCTATCGAGAGCATACCAGATCCTCCA<sup>14</sup>ACCAATATTCCAGAGGCTCTCTGGACCT  
8500 F METGLYGLYALA  
ARGARGILEARGARGSERPROGLN<sup>15</sup>ALA<sup>16</sup> ART2 (fin)  
GLYGLUPHEGLYGLUVALLEUARGLEUGLUL<sup>17</sup>EUTHRTYR<sup>18</sup>LEU<sup>19</sup>GLNTYRGLY<sup>20</sup>TRPSERTYR  
ACCGAGAA<sup>21</sup>TCAGGAGAGCTCCTCAGGCTGAACTGACCTACCTACAATATGGGTGGAGCT  
ILE SERLYS<sup>22</sup>LSARGSERLYS<sup>23</sup>PROPROGLU<sup>24</sup>I<sup>25</sup>LEYCSASPARGASP<sup>26</sup>SERCYS<sup>27</sup>GLY<sup>28</sup>ARGVAL  
PHEGLN<sup>29</sup>GLUALAV<sup>30</sup>ALGLN<sup>31</sup>ALA<sup>32</sup>ARGASP<sup>33</sup>LEU<sup>34</sup>ARGGLN<sup>35</sup>ARGLEU<sup>36</sup>LEU<sup>37</sup>ARGALA<sup>38</sup>ARGGLY  
ATTCCA<sup>39</sup>AGGA<sup>40</sup>ACGGGT<sup>41</sup>CAAGGCC<sup>42</sup>CAGAGATCTGCA<sup>43</sup>AGAGACTCTCCGGCGCGTC  
8600 GLYARGASNTYRGLY<sup>44</sup>T<sup>45</sup>ARGLEU<sup>46</sup>PHELY<sup>47</sup>SGLY<sup>48</sup>VALGLU<sup>49</sup>ASPGLY<sup>50</sup>SERSERGLN<sup>51</sup>SERLEUGLY  
GLULYS<sup>52</sup>LEUTRPG<sup>53</sup>GLUALAEUGL<sup>54</sup>NARGGLY<sup>55</sup>GLY<sup>56</sup>ARGTRP<sup>57</sup>I<sup>58</sup>LEU<sup>59</sup>ALA<sup>60</sup>ILE<sup>61</sup>PRDARGARG  
CGGAGAA<sup>62</sup>ATTATGGAGGCTCTCAAAGGGGTGAA<sup>63</sup>GAGATGGATCTCCG<sup>64</sup>AA<sup>65</sup>CTTAGGA<sup>66</sup>  
8700 GLY<sup>67</sup>LEU<sup>68</sup>ASPLYS<sup>69</sup>GLY<sup>70</sup>LEU<sup>71</sup>SE<sup>72</sup>RSER<sup>73</sup>CY<sup>74</sup>SGL<sup>75</sup>U<sup>76</sup>GLY<sup>77</sup>GLN<sup>78</sup>L<sup>79</sup>YST<sup>80</sup>R<sup>81</sup>A<sup>82</sup>S<sup>83</sup>N<sup>84</sup>GL<sup>85</sup>GLU  
ILE<sup>86</sup>ARG<sup>87</sup>GLN<sup>88</sup>GLY<sup>89</sup>LEU<sup>90</sup>GL<sup>91</sup>U<sup>92</sup>LEU<sup>93</sup>LE<sup>94</sup>U<sup>95</sup>RE<sup>96</sup>U<sup>97</sup>LE<sup>98</sup>U<sup>99</sup>LE<sup>100</sup>U<sup>101</sup>LE<sup>102</sup>U<sup>103</sup>LE<sup>104</sup>U<sup>105</sup>LE<sup>106</sup>U<sup>107</sup>LE<sup>108</sup>U<sup>109</sup>LE<sup>110</sup>U<sup>111</sup>LE<sup>112</sup>U<sup>113</sup>LE<sup>114</sup>U<sup>115</sup>LE<sup>116</sup>U<sup>117</sup>LE<sup>118</sup>U<sup>119</sup>LE<sup>120</sup>U<sup>121</sup>LE<sup>122</sup>U<sup>123</sup>LE<sup>124</sup>U<sup>125</sup>LE<sup>126</sup>U<sup>127</sup>LE<sup>128</sup>U<sup>129</sup>LE<sup>130</sup>U<sup>131</sup>LE<sup>132</sup>U<sup>133</sup>LE<sup>134</sup>U<sup>135</sup>LE<sup>136</sup>U<sup>137</sup>LE<sup>138</sup>U<sup>139</sup>LE<sup>140</sup>U<sup>141</sup>LE<sup>142</sup>U<sup>143</sup>LE<sup>144</sup>U<sup>145</sup>LE<sup>146</sup>U<sup>147</sup>LE<sup>148</sup>U<sup>149</sup>LE<sup>150</sup>U<sup>151</sup>LE<sup>152</sup>U<sup>153</sup>LE<sup>154</sup>U<sup>155</sup>LE<sup>156</sup>U<sup>157</sup>LE<sup>158</sup>U<sup>159</sup>LE<sup>160</sup>U<sup>161</sup>LE<sup>162</sup>U<sup>163</sup>LE<sup>164</sup>U<sup>165</sup>LE<sup>166</sup>U<sup>167</sup>LE<sup>168</sup>U<sup>169</sup>LE<sup>170</sup>U<sup>171</sup>LE<sup>172</sup>U<sup>173</sup>LE<sup>174</sup>U<sup>175</sup>LE<sup>176</sup>U<sup>177</sup>LE<sup>178</sup>U<sup>179</sup>LE<sup>180</sup>U<sup>181</sup>LE<sup>182</sup>U<sup>183</sup>LE<sup>184</sup>U<sup>185</sup>LE<sup>186</sup>U<sup>187</sup>LE<sup>188</sup>U<sup>189</sup>LE<sup>190</sup>U<sup>191</sup>LE<sup>192</sup>U<sup>193</sup>LE<sup>194</sup>U<sup>195</sup>LE<sup>196</sup>U<sup>197</sup>LE<sup>198</sup>U<sup>199</sup>LE<sup>200</sup>U<sup>201</sup>LE<sup>202</sup>U<sup>203</sup>LE<sup>204</sup>U<sup>205</sup>LE<sup>206</sup>U<sup>207</sup>LE<sup>208</sup>U<sup>209</sup>LE<sup>210</sup>U<sup>211</sup>LE<sup>212</sup>U<sup>213</sup>LE<sup>214</sup>U<sup>215</sup>LE<sup>216</sup>U<sup>217</sup>LE<sup>218</sup>U<sup>219</sup>LE<sup>220</sup>U<sup>221</sup>LE<sup>222</sup>U<sup>223</sup>LE<sup>224</sup>U<sup>225</sup>LE<sup>226</sup>U<sup>227</sup>LE<sup>228</sup>U<sup>229</sup>LE<sup>230</sup>U<sup>231</sup>LE<sup>232</sup>U<sup>233</sup>LE<sup>234</sup>U<sup>235</sup>LE<sup>236</sup>U<sup>237</sup>LE<sup>238</sup>U<sup>239</sup>LE<sup>240</sup>U<sup>241</sup>LE<sup>242</sup>U<sup>243</sup>LE<sup>244</sup>U<sup>245</sup>LE<sup>246</sup>U<sup>247</sup>LE<sup>248</sup>U<sup>249</sup>LE<sup>250</sup>U<sup>251</sup>LE<sup>252</sup>U<sup>253</sup>LE<sup>254</sup>U<sup>255</sup>LE<sup>256</sup>U<sup>257</sup>LE<sup>258</sup>U<sup>259</sup>LE<sup>260</sup>U<sup>261</sup>LE<sup>262</sup>U<sup>263</sup>LE<sup>264</sup>U<sup>265</sup>LE<sup>266</sup>U<sup>267</sup>LE<sup>268</sup>U<sup>269</sup>LE<sup>270</sup>U<sup>271</sup>LE<sup>272</sup>U<sup>273</sup>LE<sup>274</sup>U<sup>275</sup>LE<sup>276</sup>U<sup>277</sup>LE<sup>278</sup>U<sup>279</sup>LE<sup>280</sup>U<sup>281</sup>LE<sup>282</sup>U<sup>283</sup>LE<sup>284</sup>U<sup>285</sup>LE<sup>286</sup>U<sup>287</sup>LE<sup>288</sup>U<sup>289</sup>LE<sup>290</sup>U<sup>291</sup>LE<sup>292</sup>U<sup>293</sup>LE<sup>294</sup>U<sup>295</sup>LE<sup>296</sup>U<sup>297</sup>LE<sup>298</sup>U<sup>299</sup>LE<sup>300</sup>U<sup>301</sup>LE<sup>302</sup>U<sup>303</sup>LE<sup>304</sup>U<sup>305</sup>LE<sup>306</sup>U<sup>307</sup>LE<sup>308</sup>U<sup>309</sup>LE<sup>310</sup>U<sup>311</sup>LE<sup>312</sup>U<sup>313</sup>LE<sup>314</sup>U<sup>315</sup>LE<sup>316</sup>U<sup>317</sup>LE<sup>318</sup>U<sup>319</sup>LE<sup>320</sup>U<sup>321</sup>LE<sup>322</sup>U<sup>323</sup>LE<sup>324</sup>U<sup>325</sup>LE<sup>326</sup>U<sup>327</sup>LE<sup>328</sup>U<sup>329</sup>LE<sup>330</sup>U<sup>331</sup>LE<sup>332</sup>U<sup>333</sup>LE<sup>334</sup>U<sup>335</sup>LE<sup>336</sup>U<sup>337</sup>LE<sup>338</sup>U<sup>339</sup>LE<sup>340</sup>U<sup>341</sup>LE<sup>342</sup>U<sup>343</sup>LE<sup>344</sup>U<sup>345</sup>LE<sup>346</sup>U<sup>347</sup>LE<sup>348</sup>U<sup>349</sup>LE<sup>350</sup>U<sup>351</sup>LE<sup>352</sup>U<sup>353</sup>LE<sup>354</sup>U<sup>355</sup>LE<sup>356</sup>U<sup>357</sup>LE<sup>358</sup>U<sup>359</sup>LE<sup>360</sup>U<sup>361</sup>LE<sup>362</sup>U<sup>363</sup>LE<sup>364</sup>U<sup>365</sup>LE<sup>366</sup>U<sup>367</sup>LE<sup>368</sup>U<sup>369</sup>LE<sup>370</sup>U<sup>371</sup>LE<sup>372</sup>U<sup>373</sup>LE<sup>374</sup>U<sup>375</sup>LE<sup>376</sup>U<sup>377</sup>LE<sup>378</sup>U<sup>379</sup>LE<sup>380</sup>U<sup>381</sup>LE<sup>382</sup>U<sup>383</sup>LE<sup>384</sup>U<sup>385</sup>LE<sup>386</sup>U<sup>387</sup>LE<sup>388</sup>U<sup>389</sup>LE<sup>390</sup>U<sup>391</sup>LE<sup>392</sup>U<sup>393</sup>LE<sup>394</sup>U<sup>395</sup>LE<sup>396</sup>U<sup>397</sup>LE<sup>398</sup>U<sup>399</sup>LE<sup>400</sup>U<sup>401</sup>LE<sup>402</sup>U<sup>403</sup>LE<sup>404</sup>U<sup>405</sup>LE<sup>406</sup>U<sup>407</sup>LE<sup>408</sup>U<sup>409</sup>LE<sup>410</sup>U<sup>411</sup>LE<sup>412</sup>U<sup>413</sup>LE<sup>414</sup>U<sup>415</sup>LE<sup>416</sup>U<sup>417</sup>LE<sup>418</sup>U<sup>419</sup>LE<sup>420</sup>U<sup>421</sup>LE<sup>422</sup>U<sup>423</sup>LE<sup>424</sup>U<sup>425</sup>LE<sup>426</sup>U<sup>427</sup>LE<sup>428</sup>U<sup>429</sup>LE<sup>430</sup>U<sup>431</sup>LE<sup>432</sup>U<sup>433</sup>LE<sup>434</sup>U<sup>435</sup>LE<sup>436</sup>U<sup>437</sup>LE<sup>438</sup>U<sup>439</sup>LE<sup>440</sup>U<sup>441</sup>LE<sup>442</sup>U<sup>443</sup>LE<sup>444</sup>U<sup>445</sup>LE<sup>446</sup>U<sup>447</sup>LE<sup>448</sup>U<sup>449</sup>LE<sup>450</sup>U<sup>451</sup>LE<sup>452</sup>U<sup>453</sup>LE<sup>454</sup>U<sup>455</sup>LE<sup>456</sup>U<sup>457</sup>LE<sup>458</sup>U<sup>459</sup>LE<sup>460</sup>U<sup>461</sup>LE<sup>462</sup>U<sup>463</sup>LE<sup>464</sup>U<sup>465</sup>LE<sup>466</sup>U<sup>467</sup>LE<sup>468</sup>U<sup>469</sup>LE<sup>470</sup>U<sup>471</sup>LE<sup>472</sup>U<sup>473</sup>LE<sup>474</sup>U<sup>475</sup>LE<sup>476</sup>U<sup>477</sup>LE<sup>478</sup>U<sup>479</sup>LE<sup>480</sup>U<sup>481</sup>LE<sup>482</sup>U<sup>483</sup>LE<sup>484</sup>U<sup>485</sup>LE<sup>486</sup>U<sup>487</sup>LE<sup>488</sup>U<sup>489</sup>LE<sup>490</sup>U<sup>491</sup>LE<sup>492</sup>U<sup>493</sup>LE<sup>494</sup>U<sup>495</sup>LE<sup>496</sup>U<sup>497</sup>LE<sup>498</sup>U<sup>499</sup>LE<sup>500</sup>U<sup>501</sup>LE<sup>502</sup>U<sup>503</sup>LE<sup>504</sup>U<sup>505</sup>LE<sup>506</sup>U<sup>507</sup>LE<sup>508</sup>U<sup>509</sup>LE<sup>510</sup>U<sup>511</sup>LE<sup>512</sup>U<sup>513</sup>LE<sup>514</sup>U<sup>515</sup>LE<sup>516</sup>U<sup>517</sup>LE<sup>518</sup>U<sup>519</sup>LE<sup>520</sup>U<sup>521</sup>LE<sup>522</sup>U<sup>523</sup>LE<sup>524</sup>U<sup>525</sup>LE<sup>526</sup>U<sup>527</sup>LE<sup>528</sup>U<sup>529</sup>LE<sup>530</sup>U<sup>531</sup>LE<sup>532</sup>U<sup>533</sup>LE<sup>534</sup>U<sup>535</sup>LE<sup>536</sup>U<sup>537</sup>LE<sup>538</sup>U<sup>539</sup>LE<sup>540</sup>U<sup>541</sup>LE<sup>542</sup>U<sup>543</sup>LE<sup>544</sup>U<sup>545</sup>LE<sup>546</sup>U<sup>547</sup>LE<sup>548</sup>U<sup>549</sup>LE<sup>550</sup>U<sup>551</sup>LE<sup>552</sup>U<sup>553</sup>LE<sup>554</sup>U<sup>555</sup>LE<sup>556</sup>U<sup>557</sup>LE<sup>558</sup>U<sup>559</sup>LE<sup>560</sup>U<sup>561</sup>LE<sup>562</sup>U<sup>563</sup>LE<sup>564</sup>U<sup>565</sup>LE<sup>566</sup>U<sup>567</sup>LE<sup>568</sup>U<sup>569</sup>LE<sup>570</sup>U<sup>571</sup>LE<sup>572</sup>U<sup>573</sup>LE<sup>574</sup>U<sup>575</sup>LE<sup>576</sup>U<sup>577</sup>LE<sup>578</sup>U<sup>579</sup>LE<sup>580</sup>U<sup>581</sup>LE<sup>582</sup>U<sup>583</sup>LE<sup>584</sup>U<sup>585</sup>LE<sup>586</sup>U<sup>587</sup>LE<sup>588</sup>U<sup>589</sup>LE<sup>590</sup>U<sup>591</sup>LE<sup>592</sup>U<sup>593</sup>LE<sup>594</sup>U<sup>595</sup>LE<sup>596</sup>U<sup>597</sup>LE<sup>598</sup>U<sup>599</sup>LE<sup>600</sup>U<sup>601</sup>LE<sup>602</sup>U<sup>603</sup>LE<sup>604</sup>U<sup>605</sup>LE<sup>606</sup>U<sup>607</sup>LE<sup>608</sup>U<sup>609</sup>LE<sup>610</sup>U<sup>611</sup>LE<sup>612</sup>U<sup>613</sup>LE<sup>614</sup>U<sup>615</sup>LE<sup>616</sup>U<sup>617</sup>LE<sup>618</sup>U<sup>619</sup>LE<sup>620</sup>U<sup>621</sup>LE<sup>622</sup>U<sup>623</sup>LE<sup>624</sup>U<sup>625</sup>LE<sup>626</sup>U<sup>627</sup>LE<sup>628</sup>U<sup>629</sup>LE<sup>630</sup>U<sup>631</sup>LE<sup>632</sup>U<sup>633</sup>LE<sup>634</sup>U<sup>635</sup>LE<sup>636</sup>U<sup>637</sup>LE<sup>638</sup>U<sup>639</sup>LE<sup>640</sup>U<sup>641</sup>LE<sup>642</sup>U<sup>643</sup>LE<sup>644</sup>U<sup>645</sup>LE<sup>646</sup>U<sup>647</sup>LE<sup>648</sup>U<sup>649</sup>LE<sup>650</sup>U<sup>651</sup>LE<sup>652</sup>U<sup>653</sup>LE<sup>654</sup>U<sup>655</sup>LE<sup>656</sup>U<sup>657</sup>LE<sup>658</sup>U<sup>659</sup>LE<sup>660</sup>U<sup>661</sup>LE<sup>662</sup>U<sup>663</sup>LE<sup>664</sup>U<sup>665</sup>LE<sup>666</sup>U<sup>667</sup>LE<sup>668</sup>U<sup>669</sup>LE<sup>670</sup>U<sup>671</sup>LE<sup>672</sup>U<sup>673</sup>LE<sup>674</sup>U<sup>675</sup>LE<sup>676</sup>U<sup>677</sup>LE<sup>678</sup>U<sup>679</sup>LE<sup>680</sup>U<sup>681</sup>LE<sup>682</sup>U<sup>683</sup>LE<sup>684</sup>U<sup>685</sup>LE<sup>686</sup>U<sup>687</sup>LE<sup>688</sup>U<sup>689</sup>LE<sup>690</sup>U<sup>691</sup>LE<sup>692</sup>U<sup>693</sup>LE<sup>694</sup>U<sup>695</sup>LE<sup>696</sup>U<sup>697</sup>LE<sup>698</sup>U<sup>699</sup>LE<sup>700</sup>U<sup>701</sup>LE<sup>702</sup>U<sup>703</sup>LE<sup>704</sup>U<sup>705</sup>LE<sup>706</sup>U<sup>707</sup>LE<sup>708</sup>U<sup>709</sup>LE<sup>710</sup>U<sup>711</sup>LE<sup>712</sup>U<sup>713</sup>LE<sup>714</sup>U<sup>715</sup>LE<sup>716</sup>U<sup>717</sup>LE<sup>718</sup>U<sup>719</sup>LE<sup>720</sup>U<sup>721</sup>LE<sup>722</sup>U<sup>723</sup>LE<sup>724</sup>U<sup>725</sup>LE<sup>726</sup>U<sup>727</sup>LE<sup>728</sup>U<sup>729</sup>LE<sup>730</sup>U<sup>731</sup>LE<sup>732</sup>U<sup>733</sup>LE<sup>734</sup>U<sup>735</sup>LE<sup>736</sup>U<sup>737</sup>LE<sup>738</sup>U<sup>739</sup>LE<sup>740</sup>U<sup>741</sup>LE<sup>742</sup>U<sup>743</sup>LE<sup>744</sup>U<sup>745</sup>LE<sup>746</sup>U<sup>747</sup>LE<sup>748</sup>U<sup>749</sup>LE<sup>750</sup>U<sup>751</sup>LE<sup>752</sup>U<sup>753</sup>LE<sup>754</sup>U<sup>755</sup>LE<sup>756</sup>U<sup>757</sup>LE<sup>758</sup>U<sup>759</sup>LE<sup>760</sup>U<sup>761</sup>LE<sup>762</sup>U<sup>763</sup>LE<sup>764</sup>U<sup>765</sup>LE<sup>766</sup>U<sup>767</sup>LE<sup>768</sup>U<sup>769</sup>LE<sup>770</sup>U<sup>771</sup>LE<sup>772</sup>U<sup>773</sup>LE<sup>774</sup>U<sup>775</sup>LE<sup>776</sup>U<sup>777</sup>LE<sup>778</sup>U<sup>779</sup>LE<sup>7710</sup>U<sup>7711</sup>LE<sup>7712</sup>U<sup>7713</sup>LE<sup>7714</sup>U<sup>7715</sup>LE<sup>7716</sup>U<sup>7717</sup>LE<sup>7718</sup>U<sup>7719</sup>LE<sup>7720</sup>U<sup>7721</sup>LE<sup>7722</sup>U<sup>7723</sup>LE<sup>7724</sup>U<sup>7725</sup>LE<sup>7726</sup>U<sup>7727</sup>LE<sup>7728</sup>U<sup>7729</sup>LE<sup>7730</sup>U<sup>7731</sup>LE<sup>7732</sup>U<sup>7733</sup>LE<sup>7734</sup>U<sup>7735</sup>LE<sup>7736</sup>U<sup>7737</sup>LE<sup>7738</sup>U<sup>7739</sup>LE<sup>7740</sup>U<sup>7741</sup>LE<sup>7742</sup>U<sup>7743</sup>LE<sup>7744</sup>U<sup>7745</sup>LE<sup>7746</sup>U<sup>7747</sup>LE<sup>7748</sup>U<sup>7749</sup>LE<sup>7750</sup>U<sup>7751</sup>LE<sup>7752</sup>U<sup>7753</sup>LE<sup>7754</sup>U<sup>7755</sup>LE<sup>7756</sup>U<sup>7757</sup>LE<sup>7758</sup>U<sup>7759</sup>LE<sup>7760</sup>U<sup>7761</sup>LE<sup>7762</sup>U<sup>7763</sup>LE<sup>7764</sup>U<sup>7765</sup>LE<sup>7766</sup>U<sup>7767</sup>LE<sup>7768</sup>U<sup>7769</sup>LE<sup>7770</sup>U<sup>7771</sup>LE<sup>7772</sup>U<sup>7773</sup>LE<sup>7774</sup>U<sup>7775</sup>LE<sup>7776</sup>U<sup>7777</sup>LE<sup>7778</sup>U<sup>7779</sup>LE<sup>77710</sup>U<sup>77711</sup>LE<sup>77712</sup>U<sup>77713</sup>LE<sup>77714</sup>U<sup>77715</sup>LE<sup>77716</sup>U<sup>77717</sup>LE<sup>77718</sup>U<sup>77719</sup>LE<sup>77720</sup>U<sup>77721</sup>LE<sup>77722</sup>U<sup>77723</sup>LE<sup>77724</sup>U<sup>77725</sup>LE<sup>77726</sup>U<sup>77727</sup>LE<sup>77728</sup>U<sup>77729</sup>LE<sup>77730</sup>U<sup>77731</sup>LE<sup>77732</sup>U<sup>77733</sup>LE<sup>77734</sup>U<sup>77735</sup>LE<sup>77736</sup>U<sup>77737</sup>LE<sup>77738</sup>U<sup>77739</sup>LE<sup>77740</sup>U<sup>77741</sup>LE<sup>77742</sup>U<sup>77743</sup>LE<sup>77744</sup>U<sup>77745</sup>LE<sup>77746</sup>U<sup>77747</sup>LE<sup>77748</sup>U<sup>77749</sup>LE<sup>77750</sup>U<sup>77751</sup>LE<sup>77752</sup>U<sup>77753</sup>LE<sup>77754</sup>U<sup>77755</sup>LE<sup>77756</sup>U<sup>77757</sup>LE<sup>77758</sup>U<sup>77759</sup>LE<sup>77760</sup>U<sup>77761</sup>LE<sup>77762</sup>U<sup>77763</sup>LE<sup>77764</sup>U<sup>77765</sup>LE<sup>77766</sup>U<sup>77767</sup>LE<sup>77768</sup>U<sup>77769</sup>LE<sup>77770</sup>U<sup>77771</sup>LE<sup>77772</sup>U<sup>77773</sup>LE<sup>77774</sup>U<sup>77775</sup>LE<sup>77776</sup>U<sup>77777</sup>LE<sup>77778</sup>U<sup>77779</sup>LE<sup>777710</sup>U<sup>777711</sup>LE<sup>777712</sup>U<sup>777713</sup>LE<sup>777714</sup>U<sup>777715</sup>LE<sup>777716</sup>U<sup>777717</sup>LE<sup>777718</sup>U<sup>777719</sup>LE<sup>777720</sup>U<sup>777721</sup>LE<sup>777722</sup>U<sup>777723</sup>LE<sup>777724</sup>U<sup>777725</sup>LE<sup>777726</sup>U<sup>777727</sup>LE<sup>777728</sup>U<sup>777729</sup>LE<sup>777730</sup>U<sup>777731</sup>LE<sup>777732</sup>U<sup>777733</sup>LE<sup>777734</sup>U<sup>777735</sup>LE<sup>777736</sup>U<sup>777737</sup>LE<sup>777738</sup>U<sup>777739</sup>LE<sup>777740</sup>U<sup>777741</sup>LE<sup>777742</sup>U<sup>777743</sup>LE<sup>777744</sup>U<sup>777745</sup>LE<sup>777746</sup>U<sup>777747</sup>LE<sup>777748</sup>U<sup>777749</sup>LE<sup>777750</sup>U<sup>777751</sup>LE<sup>777752</sup>U<sup>777753</sup>LE<sup>777754</sup>U<sup>777755</sup>LE<sup>777756</sup>U<sup>777757</sup>LE<sup>777758</sup>U<sup>777759</sup>LE<sup>777760</sup>U<sup>777761</sup>LE<sup>777762</sup>U<sup>777763</sup>LE<sup>777764</sup>U<sup>777765</sup>LE<sup>777766</sup>U<sup>777767</sup>LE<sup>777768</sup>U<sup>777769</sup>LE<sup>777770</sup>U<sup>777771</sup>LE<sup>777772</sup>U<sup>777773</sup>LE<sup>777774</sup>U<sup>777775</sup>LE<sup>777776</sup>U<sup>777777</sup>LE<sup>777778</sup>U<sup>777779</sup>LE<sup>7777710</sup>U<sup>7777711</sup>LE<sup>7777712</sup>U<sup>7777713</sup>LE<sup>7777714</sup>U<sup>7777715</sup>LE<sup>7777716</sup>U<sup>7777717</sup>LE<sup>7777718</sup>U<sup>7777719</sup>LE<sup>7777720</sup>U<sup>7777721</sup>LE<sup>7777722</sup>U<sup>7777723</sup>LE<sup>7777724</sup>U<sup>7777725</sup>LE<sup>7777726</sup>U<sup>7777727</sup>LE<sup>7777728</sup>U<sup>7777729</sup>LE<sup>7777730</sup>U<sup>7777731</sup>LE<sup>7777732</sup>U<sup>7777733</sup>LE<sup>7777734</sup>U<sup>7777735</sup>LE<sup>7777736</sup>U<sup>7777737</sup>LE<sup>7777738</sup>U<sup>7777739</sup>LE<sup>7777740</sup>U<sup>7777741</sup>LE<sup>7777742</sup>U<sup>7777743</sup>LE<sup>7777744</sup>U<sup>7777745</sup>LE<sup>7777746</sup>U<sup>7777747</sup>LE<sup>7777748</sup>U<sup>7777749</sup>LE<sup>7777750</sup>U<sup>7777751</sup>LE<sup>7777752</sup>U<sup>7777753</sup>LE<sup>7777754</sup>U<sup>7777755</sup>LE<sup>7777756</sup>U<sup>7777757</sup>LE<sup>7777758</sup>U<sup>7777759</sup>LE<sup>7777760</sup>U<sup>7777761</sup>LE<sup>7777762</sup>U<sup>7777763</sup>LE<sup>7777764</sup>U<sup>7777765</sup>LE<sup>7777766</sup>U<sup>7777767</sup>LE<sup>7777768</sup>U<sup>7777769</sup>LE<sup>7777770</sup>U<sup>7777771</sup>LE<sup>7777772</sup>U<sup>7777773</sup>LE<sup>7777774</sup>U<sup>7777775</sup>LE<sup>7777776</sup>U<sup>7777777</sup>LE<sup>7777778</sup>U<sup>7777779</sup>LE<sup>77777710</sup>U<sup>77777711</sup>LE<sup>77777712</sup>U<sup>77777713</sup>LE<sup>77777714</sup>U<sup>77777715</sup>LE<sup>77777716</sup>U<sup>77777717</sup>LE<sup>77777718</sup>U<sup>77777719</sup>LE<sup>77777720</sup>U<sup>77777721</sup>LE<sup>77777722</sup>U<sup>77777723</sup>LE<sup>77777724</sup>U<sup>77777725</sup>LE<sup>77777726</sup>U<sup>77777727</sup>LE<sup>77777728</sup>U<sup>77777729</sup>LE<sup>77777730</sup>U<sup>77777731</sup>LE<sup>77777732</sup>U<sup>77777733</sup>LE<sup>77777734</sup>U<sup>77777735</sup>LE<sup>77777736</sup>U<sup>77777737</sup>LE<sup>77777738</sup>U<sup>77777739</sup>LE<sup>77777740</sup>U<sup>77777741</sup>LE<sup>77777742</sup>U<sup>77777743</sup>LE<sup>77777744</sup>U<sup>77777745</sup>LE<sup>77777746</sup>U<sup>77777747</sup>LE<sup>77777748</sup>U<sup>77777749</sup>LE<sup>77777750</sup>U<sup>77777751</sup>LE<sup>77777752</sup>U<sup>77777753</sup>LE<sup>77777754</sup>U<sup>77777755</sup>LE<sup>77777756</sup>U<sup>77777757</sup>LE<sup>77777758</sup>U<sup>77777759</sup>LE<sup>77777760</sup>U<sup>7</sup>

(fig.1B-suite 7)

FIG. 1C

séquence LTR  
 CIVET  
 versus  
 HIV-2 ROD

X	8960	8970	8980	8990	9000	9010
TCGAAGGGATTTATTACAGTCAAGAACATAGAATCTTAGACATAACTTAGAAAAGG						
:::::::	:::::::	:::::::	:::::::	:::::::	:::::::	:::::::
TGGAAGGGATGTTTACAGTGAAGAACATAAAATCTTAAATATATACTTAGAAAAGG						
X	8950	8960	8970	8980	8990	
9020	9030	9040	9050	9060		
AAGAAGGCATCATACCAGATTGGCAGATAACACTCCGGA---CCAGGAATTAGATACTAA						
::::: :: :: ::::::: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::						
AAGAAGGGATAATTGCCAGATTGGCAGAACATACACTCATGGGCCAGGAGTAAGATAACCAA						
9010	9020	9030	9040	9050		
9080	9090	9100	9110	9120		
AGATTTGGCTGGCTATGAAATTAAATCCCTGTAAATGTATCAGATGAGGCACAGGAGG						
:: ::						
TGTTCTTGGGTGGCTATGGAAGCTAGTACCACTAGATGTCCCACAAGAACGGGAGGACA						
9070	9080	9090	9100	9110		
9140	9150	9160	9170	9180		
ATGAGGAGCATTATTAGTGCACCCAGCTCAAACCTCCCAGTGGGATGACCCCTGGGAG						
:: ::						
CTGAGACTCACTGCTTAGTACATCCAGCACAAACAAGCAAGTTGATGACCCCATGGG						
9130	9140	9150	9160	9170		
9200	9210	9220	9230	9240		
AGGTTCTAGCATGAAAGTTGATCCAACCTAGCCTACACTTATGAGGCATATATTAGAT						
:: ::						
AGACACTAGTCTGGGAGTTGATCCCTTGCTGGCTTATAGTTACGAGGCTTTATTGGT						
9190	9200	9210	9220	9230		
9260	9270	9280	9290	9300		
ACCCAGAAAGAGTTGGAAGCAAGTCAGGCCTGTCAGAGAAAGAGGTTAAAGAACGGCTAG						
:: ::						
ACCCAGAGGAATTGGGCACAAGTCAGGCCTGCCAGAGGAAGACTGCAAGGCCAGACTGA						
9250	9260	9270	9280	9290		
9320	9330	9340	9350			
CCGCAAGAGGCCTTCTGAAATGGCT-GACAGGAAGGAAACT-----						
:: ::						
AAGCAAGAGGAATACCATTAGTTAAAGAACAGGAACAGCTATACTTGGTCAGGGCAGGAA						
9310	9320	9330	9340	9350		

FIG. 1C

9360 9370 9380 9390  
 ----- AGCTGAGACAGCAGGGACTTTCCACAAGGGGATCTCATG--GGGA  
 ::::::::::::::::::::: ::::::::::: ::::::: :::: : :::::  
 GTAACTAACAGAACAGCTGAGACTGCAGGGACTTCCAGAAGGGCTGTAACCAAGGGA  
 9370 9380 9390 9400 9410  
 9400 9410 9420 9430 9440 9450  
 GGTACTGGGAGGGCCGGTGGAACACCCACTTCTTGATGTATAAAATCACTGCAT  
 :: :: ::::::::::: :::: ::::::: :::: : :: ::::::::::: :: :::  
 GGGACATGGGAGGAGCTGGTGGGAACGCCCTCATATTCTCTGTATAAAATACCCGCTA  
 9430 9440 9450 9460 9470  
 9460 XX 10 20 30 40  
 TTCGCTCTGTA--TTCTGGAAGGGATTTATTACAGTGCAGAACAGACATAGAACATCTTAGAC  
 : : ::::::::::: : ::::::: ::::::: ::::::: ::::::: ::::::: :::  
 GCTTGCATTGTAACCTCTGGAAGGGATGTTTACAGTGAAGAACATAAATCTTAAAT  
 9490 XX 10 20 30 40  
 50 60 70 80 90  
 ATATACTTAGAAAAGGAAGGCATACCATACAGATTGGCAGATACACTCCGA---CCA  
 ::::::::::::::::::::: :: :: ::::::::::: :: :: :: :: :: :: :::  
 ATATACTTAGAAAAGGAAGAACGGATAATTGCAGATTGGCAGAACTCACACTCATGGGCCA  
 50 60 70 80 90 100  
 110 120 130 140 150  
 CGAATTAGATACCTAAACATGTTGGCTGGCATGGAAATTAAATCCCTGAAATGTATCA  
 :: :: ::::::: :: : :: :: :: :: :: :: :: :: :: :: :: :: :::  
 GGAGTAAGATACCCAAATGTTCTTGGGTGGCTATGGAAAGCTAGTACCAAGTAGATGTECCA  
 110 120 130 140 150 160  
 170 180 190 200 210  
 GATGAGGCACAGGGAGGATGAGGACCTTATTAGTGCACCCAGCTCAAACCTCCACTGG  
 : :: : :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :::  
 CAAGAAGGGGAGGACACTGAGACTCACTGCCTAGTACATCCAGCACAAACAAGCAAGTT  
 170 180 190 200 210 220  
 230 240 250 260 270  
 GATGACCCCTGGGAGAGGTTCTAGCATGGAACTTGTCCAACTCTAGCCTACACTTAT  
 ::::::: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :::  
 GATGACCCCGCATGGGAGGACACTGAGACTCTGGAGTTGTACCCCTGGCTTATAGTTAC  
 230 240 250 260 270 280  
 290 300 310  
 GAGGCATATATTAGATACCCAGAACAGAGTTGCAAGCA  
 :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :::  
 GAGGCTTTTATTGCG  
 290

(fig.1C-suite 1)

FIG. 2  
versus  
HIV-1.P

HIV2----- 10 20 30 40 50 env4  
 HIV2----- MHHQLLIA ILLA-SACLV YCTAVVTVVY GVPTHHEATI  
 HIV1----- HRVKEKYQHL WEGWKRWGTN LLGILHICSA TEKLWVTVYY GVPVWKEATT

HIV2----- 60 70 80 90 100 env5  
 HIV2----- PLFCATRHR- DT WG TIQCLPDHDD YOEITL-HVI EAFAWHNNTV  
 HIV1----- TLFCASDAKA YDTEVBNVVA THACVPTDPN PQEVVLVNVT ENFKMVKNDH

HIV2----- 110 120 130 140 150 env6  
 HIV2----- TEQAILEDVWH LFETSIKPCV KLTPLCVANKLGSSTESSSTGN KTTSKSTSTT  
 HIV1----- VEQNHEDIIS LWDQSLKPCV KLTPLCVSLK CTDL---GM ATNTNHSNTN

HIV2----- 160 170 180 190 200  
 HIV2----- --TTIPTDQE QEISEDTPCA RADHCSGLGE EETINCQFNH TGLERDKKKQ  
 HIV1----- S5SGENHMEK GEIX— ——HCSPNIS TSIRGKVQKE YAFFYKLDII

HIV2----- 210 220 230 240 250 env7  
 HIV2----- Y--NET-WYS KVVCETNREST NQTQCYHNNHC NTSVITESCD KHYHDAIREB  
 HIV1----- PIDNDTTSYT — ——TSC NTSVITQACP KVSFEPIPH

HIV2----- 260 env8 270 280 290 300  
 HIV2----- YCAPFGYALL RC-HDI-YWS GFAPNC SKVV ASTCTRNMET QTSTWF-GPN  
 HIV1----- YCAPAGFAIL KCNNEKTFNGT GP—CTNVS TVQCTHGIRP VVSTQLLL-H

HIV2----- 310 320 330 340 350  
 HIV2----- GTRAE—H RTYIYWGRD H-RTII-SLW KYYNLSLHCK RPGNKTVKQI  
 HIV1----- CSLAEEEVVI RSANFT—D NAKTIIIVQLH QSVE—INCI EPNNHTRESI

HIV2----- 360 370 380 390 400 env9  
 HIV2----- HLNS--GHVF ESHYQPKH—PROAHCWFKG -KVKDAMQEV KETLAKHPRY  
 HIV1----- RIQRGPGLAF VTICKIGH—MRQAHCHISR AKWHA—L KQIASKLREQ

FIG. 2

HIV2----- HIV1-----	<b>410</b> ↓ <b>420</b> <b>430</b> <b>440</b> <b>450</b> <b>RGTNDTRNIS FAAPGKGSDP EVAYMWLNCR GEFLYCNMTW FLE--WI--</b> <b>* * * * *    * * * * *    * * * * *    * * * * *</b> <b>FCHHKT-II FKQSS-GGDP EIVTHSFNCG GEFFYCNSTQ LFNSTWFNST</b>  <b>.....</b>
HIV2----- HIV1-----	<b>460</b> ↓ <b>470</b> <b>env1 480</b> <b>490</b> <b>500</b> <b>EN KTHENYAPCH IKOINTWHK VGRNVYPPR EGELSCNSTV</b> <b>* * * * *    * * * * *    * * * * *</b> <b>WSTECSNNTE GSDTITLPCR IKQFINHWQE VGKAMYAPPI SCQIRCSSNI</b>  <b>.....</b>
HIV2----- HIV1-----	<b>510</b> <b>520</b> <b>530</b> <b>540</b> <b>550</b> <b>TSIIANIDWQ NNNQTNITFS AEVAELYRL- ELDYKLV EITPIGFAPT</b> <b>* * * * *    * * * * *    * * * * *</b> <b>TGLLLTRDGG NNNNGSEIFR PGGGDHRDNW RSELYKYKVV KIEPLGVAPT</b>  <b>.....</b>
HIV2----- HIV1-----	<b>env3 560</b> <b>570</b> <b>580</b> <b>590</b> <b>600</b> <b>KEKRYSSAHG RHTRGVFVLG FLCFLATA GSAMGAAS LTVSAQSRTL</b> <b>* * * * *    * * * * *    * * * * *    * * * * *</b> <b>KAKRR--VVQ. REKRAVGI-G ALFLGFLCAA GSTMGARSMT LTVQA--RQL</b>  <b>.....</b>
HIV2----- HIV1-----	<b>610</b> <b>620</b> <b>630</b> ↓ <b>640</b> <b>env1 650</b> <b>LAGIVQQQQQ LLDVVVKRQQE LLRLTVWGTX NLQARVTAIE KYLQDQARLN</b> <b>* * * * *    * * * * *    * * * * *    * * * * *</b> <b>LSGIVQQQNH LLRAITEAQHQ LLQLTVWGIK QLQARILAVE RYLKDQQLLG</b>  <b>.....</b>
HIV2----- HIV1-----	<b>660</b> <b>670</b> <b>680</b> <b>690</b> <b>700</b> <b>SWGCCAFROVC HTTVPW VNDSLAPDWD HMTWQEWEKQ VRYLEANISK</b> <b>* * * * *    * * * * *    * * * * *    * * * * *</b> <b>IWGCSGKLIC TEAVZUHAST SNKSLEQIHN XMTWMEDRE INNYTSLINS</b>  <b>.....</b>
HIV2----- HIV1-----	↓ <b>710</b> <b>env2 720</b> <b>730</b> <b>740</b> <b>750</b> <b>SLEQAOIQQE KNMYTELQKLN SWDIFGNUD LTSWVKYIQY GVLIIVAVIA</b> <b>* * * * *    * * * * *    * * * * *</b> <b>LIEESQNQQE KHEQELLELD KWASLWNWPR ITNWLWYIKI FIMIVGGLVG</b>  <b>.....</b>

(fig.2 - suite 1)

	760	770	780	790	800
HIV2-----	LRIVIYVVQH	LSRLRKGYRP	V-FSSPPGYI	QQIHIIKKDRG	QPANEETEED
	***** *	* * * *	*	**	* **
HIV1-----	LRIVFAVLSI	VNRVRQGYSP	LSFQT	—	— HLPTPRG PDRPEGIEEE
.....					
	810	820	830	840	850
HIV2-----	GGSNCGDRYW	PWPIAYIHF	IRQLIRLLT	—	LYSIC RDILSRSLFT
	** *	*	*	***	
HIV1-----	GGERDRDRSI	ELVNGSLA-L	IWDDLRSCL	FSYHRL	— RDLLLIVIRI
.....					
	860	870	880	890	900
HIV2-----	LQLIYQNLRD	WLRLRTA-F	LQYGCEWIQE	AFQ	— AAA RATRETL —
	*	*	***	*	*
HIV1-----	VELLG--RRG	WEALKYWWNL	LQYWSQELKH	SAVSLLNATA	IAVAEGTDRV
.....					
	910	920	930	938	
HIV2-----	— AGACARG	LWRVLERIGR	CILAVPERRIR	QGAEIALL	
	**		*****	** *	**
HIV1-----	IEVVQGACRA	—	-IREIPRRIR	QGLERILL	
.....					

(fig. 2 - suite 2)

( ENV-mac  
FIG. 3 ( versus  
 / ENV-RCI

10	20	30	40	50	
MGCLGNOLLIAIC--SKCLWIICIQYVTVFYGVPAWRNATIPLFCATKNRDTWGTTOCL					
:	:::::	::	:::::::	:::::::	:::::
MM---NQLLIAILLASACLVY-CTQYVTVFYGVPTWKNATIPLFCATRNRTDWTGTLQCL					
10	20	30	40	50	
60	70	80	90	100	110
PDNDDYSELALNVTESFDAWENTVTEQAIEDVWQLFETSIKPCVKLSPLCITMRCNKSET					
:::::	:::::	::::::::::::	::::::::::::	:::::	:::
PDNDDYQEITLNVTAEFDAWNNTVTEQAIEDVWHLFETSIKPCVKLTPLCVAMKCSSTES					
60	70	80	90	100	110
120	130	140	150	160	170
DKHGLTKSSTTTASTTTTAKSYETRDIVNETS--PCVVHDNCTGLEQEPMISCKFNM					
:::	:::::	:::	:::	:::	:::
STGNNTTSK\$T--STTTTP----T-DQEQEISEDTPCARADNCGLGEEETINCQFNM					
120	130	140	150	160	
180	190	200	210	220	230
TGLKRDKKKEYNETWYSADLVCEQGNSTGNESRCYMHNCNTSVIQCCKDYWDAIRCRY					
:::	:::::	:::	:::	:::	:::
TGLERDKKQYNETWYSKDVVCETNNST-NQTQCYMHNCNTSVITESCDKHYWDAIRFRY					
170	180	190	200	210	220
240	250	260	270	280	290
CAPPGYALLRCNDTNYSGFMPNCNSKVSSCTRMMETQTSTWFRFNGTRAENRTYIYWHG					
::::::::::::::::::::	:::::::	::::::::::::	::::::::::::	::::::::::::	::::::::::::
CAPPGYALLRCNDTNYSGFAPNCNSKVASTCTRMMETQTSTWFGFNGTRAENRTYIYWHG					
230	240	250	260	270	280
300	310	320	330	340	350
RDNRTIISLNKHYNLTMKRRPGNKTVPYTMSALVFHS--OPVNERPKOAWCRFGGNW					
::::::::::::	:::	:::::::	:::	:::	:::
RDNRTIISLNKYNNLSLHKRPGNKTVKOJMLMSGHVFHSHYQPIKRPROAWCFKGKW					
290	300	310	320	330	340
360	370	380	390	400	
KEAIKEVKOTIVKHPRYTGTNNNTDKINLTAPRGG-DPEVTFMHTNCRGEFLYCKMNWFLN					
::	:::::	:::::	:::	:::::	:::::
KDAMQEVKETLAKHPRYRGTNDTRNISFAAPGKGSDPEVAYNNTNCRGEFLYCNMTWFLN					
350	360	370	380	390	400

FIG. 3

420 430 440 450 460  
 KVEDRSLTOKPKERHKRNYVPCHIQIINTWHKVGKNVYLPPREGDLTCNSTVTSLIAN  
 :  
 WIEN-----KT-H-RNYAPCHIKQIINTWHKVGGRNVYLPPRECELSCNSTVTSLIAN  
 410 420 430 440 450  
  
 480 490 500 510 520  
 INWHTDGNQTSITMSAEVAELYRLELGDKLVEITPIGLAPTNVKRYTTG-GTSRNKRGVF  
 :  
 IDHQNQQNQTNITFSAEVAELYRLELGDKLVEITPIGAPTKEKRYSSAHG--RHTRGVF  
 460 470 480 490 500 510  
  
 540 550 560 570 580  
 VLGFGLGFLATAGSAMGAASLTVAQSRTLLAGIVQQQQQLLDVVKRQOELLRLTVWGTKN  
 ::::::::::::::::::::: ::::::::::::::::::::: :::::::::::::::::::::  
 VLGFGLGFLATAGSAMGAASLTVAQSRTLLAGIVQQQQQLLDVVKRQOELLRLTVWGTKN  
 520 530 540 550 560 570  
  
 600 610 620 630 640  
 LOTRVSAIEKYLKDQAOLNNAWGCARQYCHTTVPWPNASLTPDNNETWQEWERKVDFLE  
 :  
 LQARYTAIEKYLQDQARLNNSWGCARQYCHTTVPWVNDLAPDWDNMKTWOEWEKQVRYLE  
 580 590 600 610 620 630  
  
 660 670 680 690 700  
 ANITALLEEAOIQQEKNMYELOKLNSHDVFGNWFDLTSWIKYIQQGIYIIIVGVILLRIVI  
 :  
 ANISKSLEAQAOIQQEKNMYELOKLNSHOIFGNWFDTSHVVKYIQQGVLIIVAVIALRIVI  
 640 650 660 670 680 690  
  
 720 730 740 750 760  
 YIVOMLARLRQGYRPVFSSPPSYFQ\*THTODPALPTKEGKKGDGGSGGNSSWPWHOIEY  
 :  
 YVVOHLSRLRKGYRPVFSSPPGYIQQIHMKDRGQANEETEEDGGSNGGDRYWPWPIAY  
 700 710 720 730 740 750  
  
 780 790 800 810 820  
 IHFLIRQLIRLLTWFNSCRTLLSRAYQILOPQFORLSATYGEFGEVLRELTYLQYGWS  
 :  
 IHFLIRQLIRLLTRLYSICRDLLSRSFLTOLIYONLRDW-----LRLRTAFLQYGCE  
 760 770 780 790 800  
  
 840 850 860 870 880  
 YFOEAQAA-RDLRQRLLRA-RGEKLWEALORGGRWILAIPRRIROGLELTL  
 :  
 WIOEAFOAAARATRETLAGACRG--LWRVLERIGRGLAYPRRIROGAEIALL  
 810 820 830 840 850

(fig. 3-suite 1)

FIG. 4 (GAG-mac  
(versus  
(GAG-ROD

10	20	30	40	50	
VQHKKEIAVFYPGRDNKIEWEMGARNSVLSGKKADELEKIRLRPGGKKYMLKHVVWAAN					
::: ::::: ::::: ::::: ; :: :: ::					
MGARNSVLRGKKADELERIRLRPGGKKYRLKHIVWAAN					
10	20	30			
70	80	90	100	110	
ELDRFGLAESLLENKEGCQKILSVLAPLVPTGSENLKSLYNTVCVIWCIAEAKVKHTEE					
::: ::::: ::::: :: : ::::: ::::: ::::: ::::: ::					
KLDRFGLAESLLESKEGCQKILTVLDPMVPTGSENLKSLFNTVCVIWCIAEAKVKDTEG					
40	50	60	70	80	90
130	140	150	160	170	
AKOIVQRHLVYMETGTAETMPKTSRPTAPFSGRGGNYPVQQIGGNYTHLPLSPRTLNAWVK					
::: ::::: ::::: :: : ::::: ::::: ::::: ::::: ::					
AKOIVRRHLVAETGTAEKMPSTSRTAPSSEKGCGNYPVQHVCGNYTHIPLSPRTLNAWVK					
100	110	120	130	140	150
190	200	210	220	230	
LIEEKKFGAEVYSGFQALSEGCLPYDINQMLNCVGDHQAMQIIRDIIINEAAADWDLQHP					
: ::::::: ::::::: ::::::: ::::::: ::::::: :::: ::					
LVEEKKFGAEVYVPGFQALSEGCTPYDINQMLNCVGDHQAMQIREIIINEAAEWDVQHP					
160	170	180	190	200	210
250	260	270	280	290	
QQAPQQ-GQLREPSSDIAGTTSTVEEQIQWMYRQONPIPVGNIYRRWIOLGLQKCVRMY					
: ::::::: ::::::: ::::::: ::::::: ::::::: ::::::: ::					
IPGPLPAGQLREPRGSDIAGTTSTVEEQIQWFRPONPVPGNIYRRWIOLGLQKCVRMY					
220	230	240	250	260	270
300	310	320	330	340	350
NPTNILDVKQGPKEPFQSYVDRFYKSLRAEQTDPAVKNWMTOTLLIONANPDCKLVLKGL					
: ::::::: ::::::: ::::::: ::::::: ::::::: ::::::: ::					
NPTNILDIKQGPKEPFQSYVDRFYKSLRAEQTDPAVKNWMTOTLLVQNANPDCKLVLKGL					
280	290	300	310	320	330
360	370	380	390	400	410
GTNPTELLEMKTACQGVGGPGQKARLMAEALKEALAPAPIFAAAQQKGPRKPIKCHNCGK					
: ::::::: ::::::: ::::::: ::::::: ::::::: :: :: ::					
GMNPTELLEMKTACQGVGGPGQKARLMAEALKEVIGPAPIFAAAQQ---RKAFKCHNCGK					
340	350	360	370	380	390

FIG. 4

420        430        440        450        460        470  
EGHSAROCRAPRROGCKCGKMDHVMAKCPNROAGFLGLGPWKKPRNFPMAQVHQLTP  
::::::::::: ::::: :: ::::::::::::: ::::: :::::  
EGHSAROCRAPRROGCKCGKPGHIMTNCPDROAGFLGLGPWKKPRNFPVAQVPOGLTP  
400        410        420        430        440        450  
  
480        490        500        510  
TAPPEEPAVDLLKNYMHLGKQQRESRGKPYKEVTEDLLHL-----NS  
::::: ::::::: :: :: :: : ::::::::::::: :::  
TAPPVDPAVDLLEKYMHQGKRQREQRERPYKEVTEDLLHLEQGETPYREPPTEDLLHLNS  
460        470        480        490        500        510

(fig.4 - suite 1)

( POL-mac  
( versus  
( POL-ROD

FIG. 5

410 420 430 440 450 460  
 DRVVLQLKELLNSIGFSSPEEKFKDKDPPFWHMGYELWPTKWKLOKIELPORETWTVDNIO  
 ::::::::::::::::::::: :::: : ::::::: ::::::::::::::::::::: :::: : :::::::  
 DRVVLQLKELLNLGLFSTPDEKFQKDPPYHWMGYELWPTKWKLOKIQLOPOKEIWTVDNIO  
 390 400 410 420 430 440  
  
 470 480 490 500 510 520  
 KLVGVLNHAAQIYPGIKTKHLCLIRGKMTLTEEVONTMAEAEYEENKIJL5QE0EGCY  
 ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: :::: : :::::::  
 KLVGVLNHAAOLYPGIKTKHLCLIRGKMTLTEEVONTELAEEALEENRILS5QE0EGHY  
 450 460 470 480 490 500  
  
 530 540 550 560 570 580  
 YOESKPLEATVIKSODNQWSYKIHOEDKILKGKFAKIKNTHTNGVRLAHVIOKICKEA  
 :::: : ::::: : : ::::: ::::::: ::::::: ::::: ::::::: ::::: : :::::::  
 YOEKELEATVKDQENOWTYKIHOEELKILKGKYAKVKNTHTNGIRLLAOVVOKICKEA  
 510 520 530 540 550 560  
  
 590 600 610 620 630 640  
 [Y]WGCVPKFHLPYEWQHNTDYWQVTWIPETPPLVRLVFNLVKDPIEGEET  
 ::::: ::::::: ::::: ::::::: ::::::: ::::::: ::::::: ::::::: ::::::: :::  
 LYIWGRIPKFHLPYVEREITWEGHWDNTYWQVTWIPDMDFYSTPPLVRLAFNLVGDPIPAGET  
 570 580 590 600 610 620  
  
 650 660 670 680 690 700  
 YYVDGCSKOSKECKAGYITDRGKDKVYLEQTTNOQAELAFLHALTDSGPKANIIYDS  
 ::::: ::::::: ::::::: ::::::: ::::::: ::::::: ::::::: ::::::: :::::::  
 FYTDGSCNRROSKECKAGYVYDTRGKDKVKKLEQTTNOQAELAFLHALTDSGPKVNIYDS  
 630 640 650 660 670 680  
  
 710 720 730 740 750 760  
 QYVMGIITGCPTESESRLVNQIIIEEMIKKTEIYYAWVPAHKGIGGNGEIDHLYSOGIROV  
 ::::: ::::::: ::::::: ::::::: ::::::: ::::::: ::::::: ::::::: :::::::  
 QYVMGISAQPTESESKIVNQIIIEEMIKKEAIYYAWVPAHKGIGGNGEVDHLYSOGIROV  
 690 700 710 720 730 740  
  
 770 780 790 800 810 820  
 LFLEKIEPAQEEHSKYHSNIKELYVFKGLPRLYAKQIVDTCDKCHOKGEAIHGQVNSDLG  
 ::::::: ::::::: ::::::: ::::::: ::::::: ::::::: ::::::: ::::::: :::  
 LFLEKIEPAQEEHEKYHSNVKELSHKFGIPMLYARQIVNSCADCOOKGEAIHGQVNAELG  
 750 760 770 780 790 800  
  
 830 840 850 860 870 880  
 THQMDCTHLEGKIVIVAVHVASGFIEAEVIPQETGROTALFLLKLASRWPITHLHTDNGA  
 ::::::: ::::::: ::::::: ::::::: ::::::: ::::::: ::::::: ::::::: :::::::  
 THQMDCTHLEGKIIIIVAVHVASGFIEAEVIPQESGROTALFLLKLASRWPITHLHTDNGA  
 810 820 830 840 850 860  
  
 890 900 910 920 930 940  
 NFASQEVKHYAWHAGIEHTFGVPYNPQSOGVYEAHHHLKNQIDRIREQANSVETIVLMA  
 ::::::: ::::::: ::::::: ::::::: ::::::: ::::::: ::::::: ::::::: :::::::  
 NFTSQEVKHYAWHIGIEQSFGVPYNPQSOGVYEAHHHLKNQISRIREQANTJETIVLMA  
 870 880 890 900 910 920  
  
 950 960 970 980 990  
 VHCHNFKRRGGIGDMTPAERLINMITTEQEIQFOOSKNSKFKNFRVYYREGRDOLWKPG  
 ::::::: ::::::: ::::::: ::::::: ::::::: ::::::: ::::::: ::::::: :::::::  
 IHCHNFKRRGGIGDMTPSERLINMITTEQEIQFOLOAKNSKLDFRYYREGRDOLWKPG  
 930 940 950 960 970 980  
  
 1010 1020 1030 -1040 1050  
 ELLWKGECAVILKVGTDIKVYPRRKAKIIKDYGGCKEMOSSHMEDTGEAREVA

(fig.5-suite 1)

**EP 0 750 041 A2**

::::::: ::::::: " ::::::: ::::: :::: :: : : :  
ELLWKGEHAVLVKYGTDIKIIPRRKAKIIRDYGGROEMDSCSHLEGAREDGEMA  
990 1000 1010 1020 1030

(fig. 5-suite 2)

FIG. 6 (Q.mac  
 versus  
 (Q.ROD

10	20	30	40	50
MEEEKRWIVVPTWRIPERLERWHSLIKYLKYKTKDLOKACYVPHHKVGWAWHTCSRVIFF				
::: :::::::::: : : : ::::: :::::::::: : :::::::::::::::::::::				
MEEDKRWIVVPTWRVPGRMEKWHSLVKYLYKTKDLEKVCYVPHHKVGWAWHTCSRVIFF				
10	20	30	40	50
70	80	90	100	110
LQEGSHLEVQGYWNLTPERGWLSTYAVRITWYSKDFWTDVTPEYADILLHSTYFPCFTAG				
: ::::: : ::::::: ::::: : ::::::: ::::::: : :: : :::::::::::::				
LKGNSHLEIQAYWNLTPEKGWLSSYSVRITWYTEKFWTDVTPDCADVLIHSTYFPCFTAG				
70	80	90	100	110
130	140	150	160	170
EVRRAIRGERLLSCCRFPRAHKHQVPSLOYLALRYVSHV-RSQGENPTWKQWRRDNRRSL				
: ::::::: ::::: ::::::: ::::: :: : : : :: : :				
EVRRAIRGEKLLSCNYPRAHRAQVPSLQFLALVVVQONDPRQDSTTRKQRRRDYRRGL				
130	140	150	160	170
180	190	200	210	
RYAKQNSRGDKQRGKGPPTEGANFPGLAKVLGILA				
: :: : :: : : : :: : :: : :: :				
RLAKQDSRSHKQRSSESPTPRTYFPGVAEVLEILA				
190	200	210		

FIG. 6

( R.mac  
FIG. 7 ( versus  
          ( R.ROD

ME---ERPPENEGPOREPWEVYVEYLKELKEEALKHFDPRLLTALGNHIYNRHGOTLE  
10       20       30       40       50  
:   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :  
MAEAPTELPPVTDGTPLREPGDEWIIEILREIKEALKHFDPRLLIALGKYIYTRHGOTLE  
10       20       30       40       50  
  
GAGELIRILQRALFIHFRSGCSHSRIGQPGGNPLSTIPPSRSML  
60       70       80       90       100  
:   :   :   :   :   :   :   :   :   :   :   :   :   :  
GARELIKVLQRALFTHFRAGCGHSRIGQTRGGNPLSAIPTPRNMQ  
70       80       90       100

FIG. 7

FIG. 8 ( X.mac  
          ( versus  
          ( X.ROD

10	20	30	40	50
MSDPRERIPPGNSGEETIGEAEFLNRTVEEINREAVNHLPRELIFOVWORSWEYWHDEQ	:: ::::: ::::::::::::::: ::::::: ::::::::::::::::::::: :::::::	MTOPRETVPNGNSGEETIGEAFALNRTVEAINREAVNHLPRELIFOVWORSWRYWHDEQ		
10	20	30	40	50
70	80	90	100	110
GMSOSYTKYRYLCLIQKALFMHCKKGCRCLGEGHGAGGGWRPGPPPPPPPLGA	::: ::::::: ::::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::	GMSESYTKYRYLCIIQAKAVYMHVRKGCTCLGRGHGPCCWRPGPPPPPPGLV		
70	80	90	100	110

FIG. 9 ( F.mac  
 ( versus  
 ( F.ROD

FIG. 9

FIG.10 ( TAT.mac  
                ( versus  
                ( TAT.ROD

10	20	30	40	50
METPLREQENSLESSNERSYISEAAAAIPESANLGEELSQLYRPLEACYNTCYCKKKC ::::: : :: : :: : :: : : : ::::::::::::: : : :::: : :: :: ::				
METPLKAPESSLKSCNEPFSRTSEQDVATOELARQGEEILSQLYRPLECNNSCYCKRCC 10 20 30 40 50				
70 80 90 100 110				
YHCQFCFLKKGLGISYEKSHRRRTPKKAKANTSSASNERP---IPNRIRLCOPKKAKKE ::::: :::: ::::: :: : ::::::: : : : : : : : :: : :: :				
YHCQMCFLNKGLGICYERKGRRRTPKKTKTHPSPT---PDKSISTRTGDSOPTKKQKK 70 80 90 100 110				
120 130				
TVEAAVATAPGLGR ::: : : : ::				
TVEATVETDTCPGR 120 130				

FIG. 10

FIG.11 { ART.mac  
          { versus  
          { ART.ROD

10	20	30	40	50	
MRSHTGEEELRRRLRLIHLHOTSKYGLSHKSAAYRHLLVDPYPTGSGSANORRQKRRW					
:	::	:::	::::	::::	
MNERADEEGLQRKLRLIRLLHQTN-----				PYPQGPCTASQRNRNRRRW	
10	20	30	40		
70	80	90	100	110	
RORWOOLLALADRIYSFPDPPTDTPLDLAIQQLONLAIESIPDPPTNIPEALCDLRRIRR					
::: :	:::::	::	:::::	::	
KORWROLALADASIYTFPDPPADSPLDQTHLQLGLTQELPDPPTHLPSQRLAET					
50	60	70	80	90	100

SPOA

**FIG. 11**